SPECIAL COLLECTOR'S EDITION

SCIENTIFIC AMERICAN

REVOLUTIONS IN SCIENCE

INCLUDING

ORIGINS OF LIFE OVERTURNED

DINOSAURS' TRUE COLORS

RETHINKING WEIGHT LOSS

— PLUS —

3

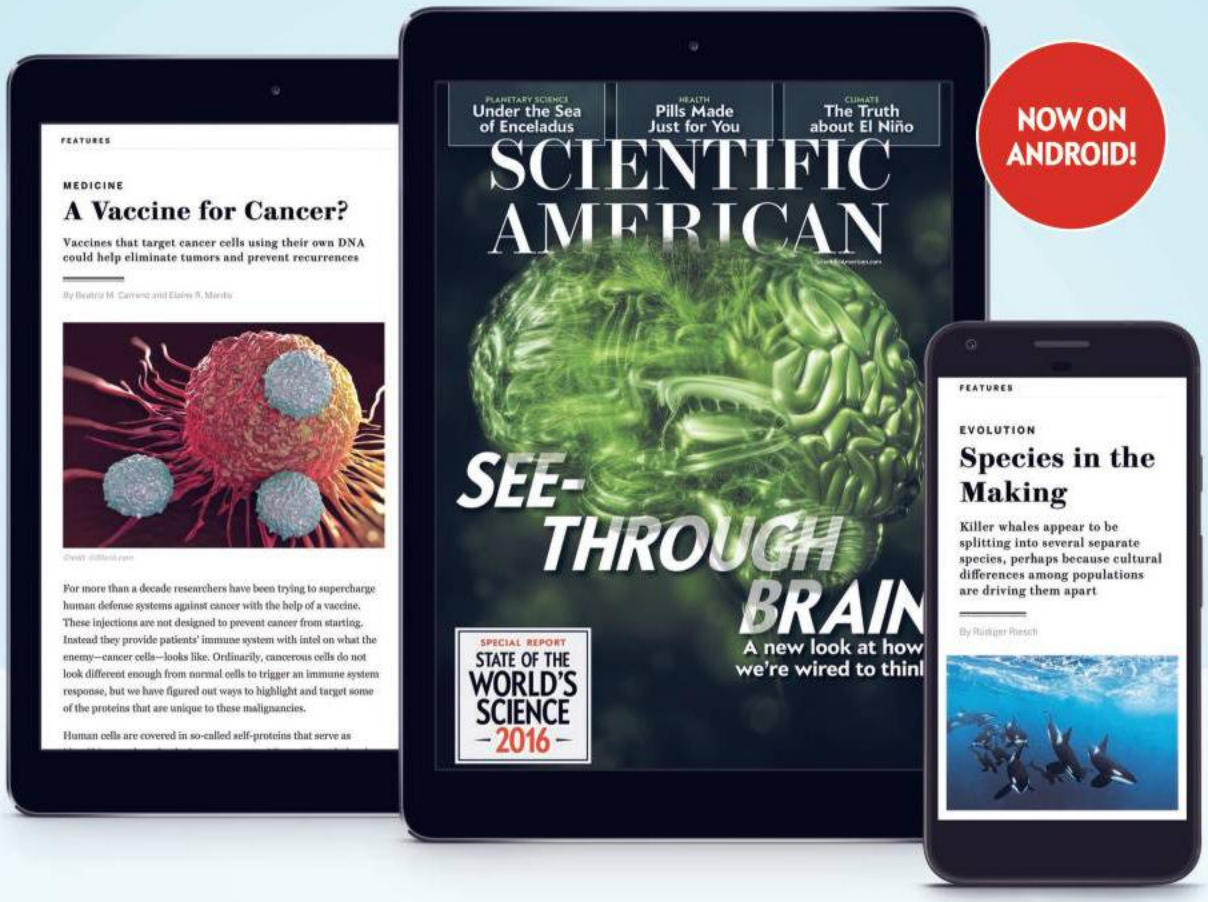
UNSOLVED SCIENTIFIC MYSTERIES

13 DISCOVERIES THAT COULD CHANGE EVERYTHING

FALL 2018

© 2018 Scientific American

Access *Scientific American* anytime, anywhere.


FEATURES
MEDICINE
A Vaccine for Cancer?
Vaccines that target cancer cells using their own DNA could help eliminate tumors and prevent recurrences
PLANETARY SCIENCE
Under the Sea of Enceladus
HEALTH
Pills Made Just for You
CLIMATE
The Truth about El Niño
SCIENTIFIC AMERICAN
SEE-THROUGH BRAIN
A new look at how we're wired to think
SPECIAL REPORT
STATE OF THE WORLD'S SCIENCE
2016
NOW ON ANDROID!
FEATURES
EVOLUTION
Species in the Making
Killer whales appear to be splitting into several separate species, perhaps because cultural differences among populations are driving them apart

Download the App for iOS™ and Android™

Download on the App Store

GET IT ON Google Play

Full Content
Read the latest issue on your phone or tablet—online or offline.

Easy Navigation
Browse articles and past issues quickly with thumbnail views.

Resize Text
Every article customized to your liking—on phone or tablet screens.

Scrapbook
Save your favorite articles to your scrapbook for future reference.

Done Reading?
Remove issues easily with the Archive button. You can re-download anytime.

Apple and the Apple logo are trademarks of Apple Inc., registered in the U.S. and other countries. App Store is a service mark of Apple Inc. Android and the Google Play logo are trademarks of Google Inc. IOS is a trademark or registered trademark of Cisco in the U.S. and other countries and is used under license.

SCIENTIFIC AMERICAN
ESTABLISHED 1845

Revolutions in Science is published by the staff of SCIENTIFIC AMERICAN, with project management by:

EDITOR IN CHIEF AND SENIOR VICE PRESIDENT
Mariette DiChristina
MANAGING EDITOR **Curtis Brainard**
COPY DIRECTOR **Maria-Christina Keller**
CREATIVE DIRECTOR **Michael Mrak**

EDITORIAL
COLLECTIONS EDITOR
Andrea Gawrylewski

ART
ISSUE DESIGNER **Lawrence R. Gendron**
SENIOR GRAPHICS EDITOR **Jen Christiansen**
PHOTOGRAPHY EDITOR **Monica Bradley**
ASSISTANT PHOTO EDITOR **Liz Tormes**

COPY AND PRODUCTION
SENIOR COPY EDITOR **Daniel C. Schlenoff**
COPY EDITOR **Aaron Shattuck**
MANAGING PRODUCTION EDITOR **Richard Hunt**
PREPRESS AND QUALITY MANAGER **Silvia De Santis**
EDITORIAL ADMINISTRATOR **Ericka Skirpan**
SENIOR SECRETARY **Maya Harty**

DIGITAL
TECHNICAL LEAD **Nicholas Sollecito**
SENIOR WEB PRODUCER **Ian Kelly**
WEB PRODUCER **Jessica Ramirez**

PRESIDENT
Dean Sanderson
EXECUTIVE VICE PRESIDENT **Michael Florek**
EXECUTIVE VICE PRESIDENT, GLOBAL ADVERTISING AND SPONSORSHIP **Jack Laschever**
PUBLISHER AND VICE PRESIDENT **Jeremy A. Abbate**

MARKETING AND BUSINESS DEVELOPMENT
HEAD, MARKETING AND PRODUCT MANAGEMENT **Richard Zinken**
MARKETING DIRECTOR, INSTITUTIONAL PARTNERSHIPS AND CUSTOMER DEVELOPMENT **Jessica Cole**
ONLINE MARKETING PRODUCT MANAGER **Zoya Lysak**

INTEGRATED MEDIA SALES
DIRECTOR, INTEGRATED MEDIA **Jay Berfas**
DIRECTOR, INTEGRATED MEDIA **Matt Bondlow**
SENIOR ADMINISTRATOR, EXECUTIVE SERVICES **May Jung**

CONSUMER MARKETING
DIGITAL MARKETING MANAGER **Marie Cummings**
E-MAIL MARKETING MANAGER **Chris Monello**
MARKETING AND CUSTOMER SERVICE COORDINATOR **Christine Kaelin**

ANCILLARY PRODUCTS
ASSOCIATE VICE PRESIDENT, BUSINESS DEVELOPMENT **Diane McGarvey**
CUSTOM PUBLISHING EDITOR **Lisa Pallatroni**
RIGHTS AND PERMISSIONS MANAGER **Felicia Ruocco**

CORPORATE
HEAD, COMMUNICATIONS, USA **Rachel Scheer**
COMMUNICATIONS MANAGER **Lauren Kuhn**

PRINT PRODUCTION
ADVERTISING PRODUCTION CONTROLLER **Carl Cherebin**
PRODUCTION CONTROLLER **Madelyn Keyes-Milch**

COVER IMAGE BY SCIENTIFIC AMERICAN; THIS PAGE: ALAMY

FROM THE EDITOR

Science, Upended

NORMALLY SCIENCE FOLLOWS AN ORDERLY PATH: NEW OBSERVATIONS BUILD on previous findings, often in incremental steps over many years, contributing piecemeal to a larger body of knowledge. Every now and then, insights arise that don't quite jibe with the observations that came before. Such aberrations yield a more nuanced understanding of the topic and are folded into the narrative of what we know. Sometimes, though, a new discovery is so profound that it is impossible to reconcile, and knowledge must proceed in a different direction entirely. This is a collection of articles about *those* kinds of advances—what Thomas S. Kuhn called "paradigm shifts" in his 1962 book *The Structure of Scientific Revolutions*.

Take, for instance, the origin of life. For more than 40 years the hypothesis has been that the most basic fragments of life assembled into more complex organisms while floating in the primordial oceans. Yet recent evidence suggests that nascence may have instead happened in the toxic pools of an active volcanic landscape (*page 4*). Or consider the dinosaurs, a legion of what researchers once believed were impressive but drab-looking creatures that dominated the earth for nearly 150 million years. In fact, they weren't drab at all—research on fossilized melanin cells now shows that dinosaurs were feathered in bright colors; some even sported dots and stripes (*page 12*).

The Bessemer process revolutionized the production of cheap steel in the late 1800s—allowing its use to expand exponentially.

Revolutionary discoveries, though uncommon, have happened in nearly every field. In health, exercise, contrary to the received wisdom, does not control body weight (*page 58*). In brain science, researchers are abandoning Noam Chomsky's long-standing theory that children's minds are hardwired for language development (*page 90*). In cosmology, the latest studies show that our solar system may not have formed by neat and tidy accretion but rather through violent collisions (*page 44*). In biology, a naturalist helped to overturn a 150-year-old understanding of lichens, shedding fresh light on how symbioses work and how life-forms are defined (*page 28*). Some of these paradigm shifts are by no means complete, but they have all forced scientists to question assumptions and adopt new perspectives.

One of my graduate school advisers used to say that what humans Know (capital "K") is infinitesimal compared with what is knowable in the universe. In that spirit, we also present a small sample of some of the most fascinating unanswered questions, from the befuddling mystery of what actually is dark matter (*page 96*) to whether quantum computing might leave traditional computers in the dust (*page 108*).

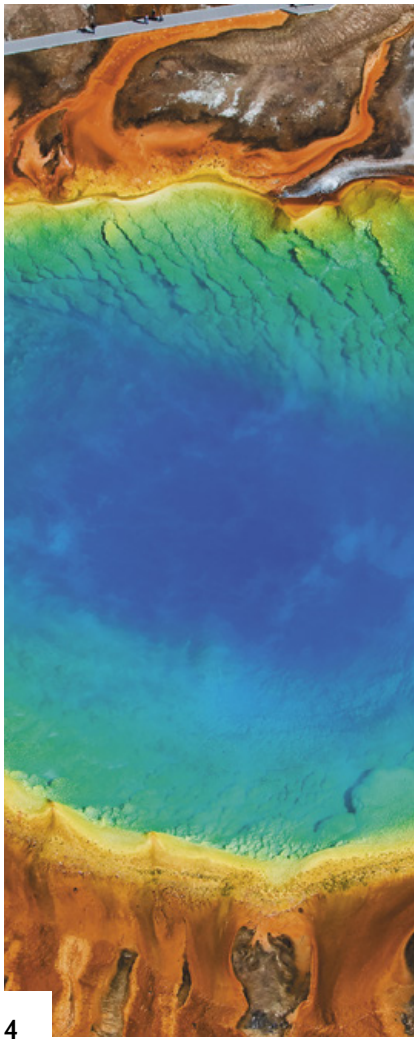
The beauty of inquiry is that when we upend decades—or longer—of thinking, we uncover an entirely new set of questions to ask. The work of science is never done. SA

Andrea Gawrylewski
Collections Editor
editors@sciam.com

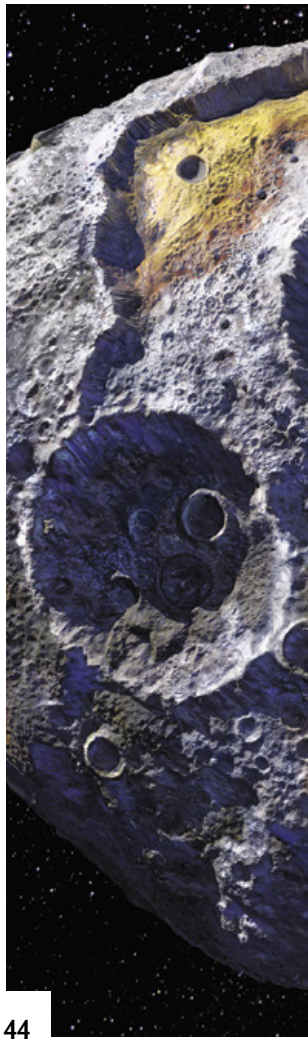
© 2018 Scientific American

SCIENTIFIC AMERICAN

Volume 27, Number 3, Fall 2018



4



44

SPECIAL EDITION

Revolutions in Science

■ *BIOLOGY & EVOLUTION*

4 Life Springs
Deep oceans were thought to hold life's origins. New evidence points instead to an active volcanic landscape. *By Martin J. Van Kranendonk, David W. Deamer and Tara Djokic*

12 The True Colors of Dinosaurs
Long thought impossible, preservation of fossil pigments is allowing scientists to reconstruct extinct organisms with unprecedented accuracy—a feat that is yielding surprising insights into the lives they led. *By Jakob Vinther*

20 The New Origins of Technology
Ancient stone tools from Kenya shatter the classic story of when and how humans became innovators. *By Kate Wong*

28 The Meaning of Lichen
How a naturalist's observations in the wilds of British Columbia inspired a scientist to discover hidden symbioses—overturning 150 years of accepted scientific wisdom. *By Erica Gies*

36 The Perfect Beast
Mixotrophs, tiny sea creatures that hunt like animals but grow like plants, can change everything from fish populations to rates of global warming. *By Aditee Mitra*

■ *COSMOS*

44 Solar System Smashup
Our neighborhood of planets was not created slowly, as scientists once thought, but in a speedy blur of high-energy crashes, destruction and rebuilding. *By Linda T. Elkins-Tanton*

52 Black Holes from the Beginning of Time
A hidden population of black holes born less than one second after the big bang could solve the mystery of dark matter. *By Juan García-Bellido and Sébastien Clesse*

■ *HEALTH*

58 The Exercise Paradox
Studies of how the human engine burns calories help to explain why physical activity does little to control weight—contrary to years of common wisdom. *By Herman Pontzer*

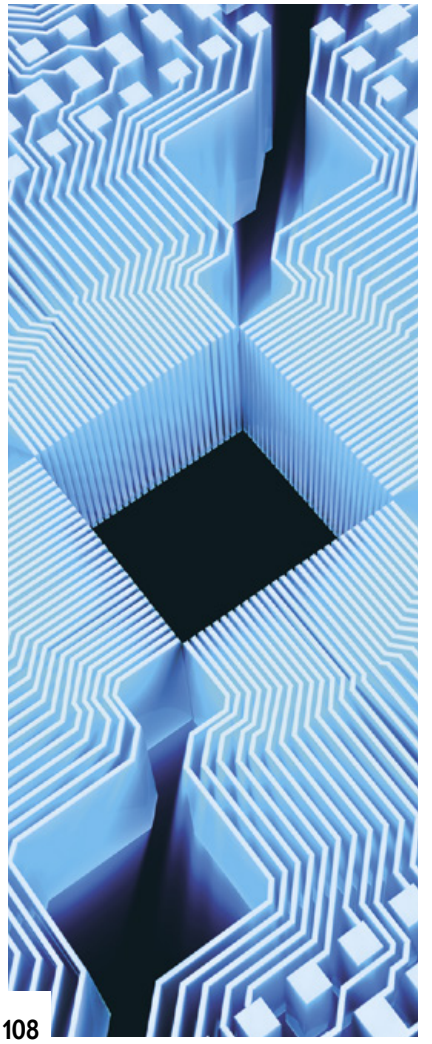
64 The New Migraine Meds
The cause of migraine headaches has eluded scientists for centuries. Now a theory blaming one nerve has led to drugs that prevent attacks. *By David Noonan*

68 Ebola's Second Coming
Brain deficits and more torment many virus survivors in Liberia. Even after clear blood tests, the virus can hide in reservoirs throughout the body. *By Seema Yasmin*

© 2018 Scientific American

68

82



108

■ *MIND*

74 The Brain, Reimagined
Physicists who have revived experiments from 50 years ago say nerve cells communicate with mechanical pulses, not electric ones. *By Douglas Fox*

82 Memory's Intricate Web
A technical revolution provides insight into how the brain links memories, a process critical for understanding and organizing the world around us. *By Alcino J. Silva*

90 Language in a New Key
Much of Noam Chomsky's paradigm of linguistics—including its account of the way we learn languages—is being overturned. *By Paul Ibbotson and Michael Tomasello*

■ *UNSOLVED QUESTIONS*

96 The Puzzle of Dark Energy
Why is the expansion of the universe accelerating? After two decades of study, the answer is as mysterious as ever, but the questions have become clearer. *By Adam G. Riess and Mario Livio*

102 The Red Line
Scientists are on the threshold of crossing a sharp ethical line: permanently altering the human genetic code. *By Stephen S. Hall*

108 Quantum Computers Compete for "Supremacy"
Two technologies may be on the verge of surpassing even the most powerful digital computers in a year or so, but key challenges remain unsolved. *By Neil Savage*

DEPARTMENTS

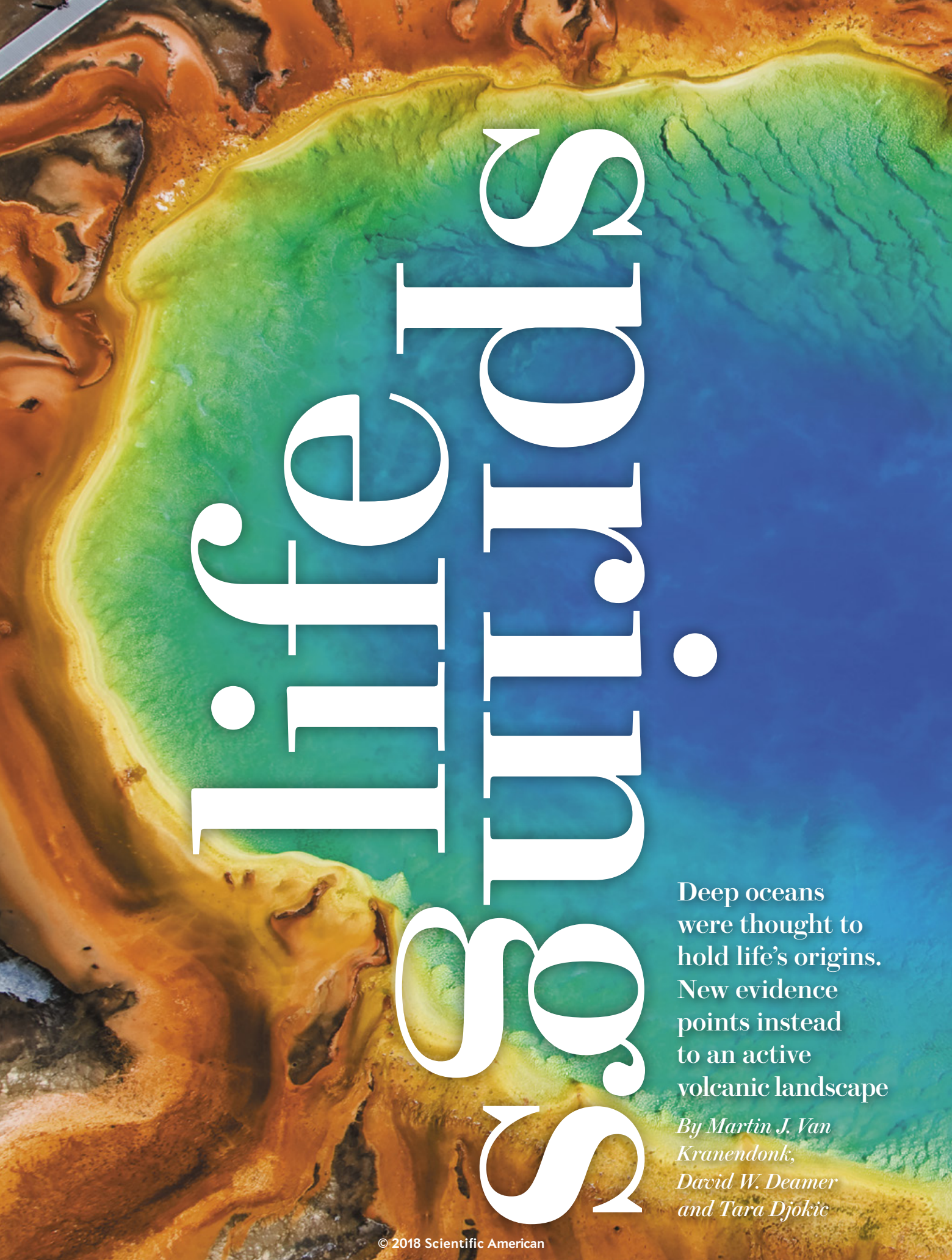
1 FROM THE EDITOR
Science, Upended

112 END NOTE
Sleeping Beauties of Science
Some of the best research can slumber for years. *By Amber Williams*

Scientific American Special (ISSN 1936-1513), Volume 27, Number 3, Fall 2018, published by Scientific American, a division of Nature America, Inc., 1 New York Plaza, Suite 4500, New York, NY 10004-1562. Canadian BN No. 127387652RT; TVQ1218059275 TQ0001. To purchase additional quantities: U.S., $13.95 each; elsewhere, $17.95 each. Send payment to Scientific American Back Issues, 1 New York Plaza, Suite 4500, New York, NY 10004-1562. **Inquiries: fax 212-355-0408 or telephone 212-451-8415. Printed in U.S.A.**

Copyright © 2018 Scientific American, a division of Nature America, Inc. All rights reserved.

© 2018 Scientific American

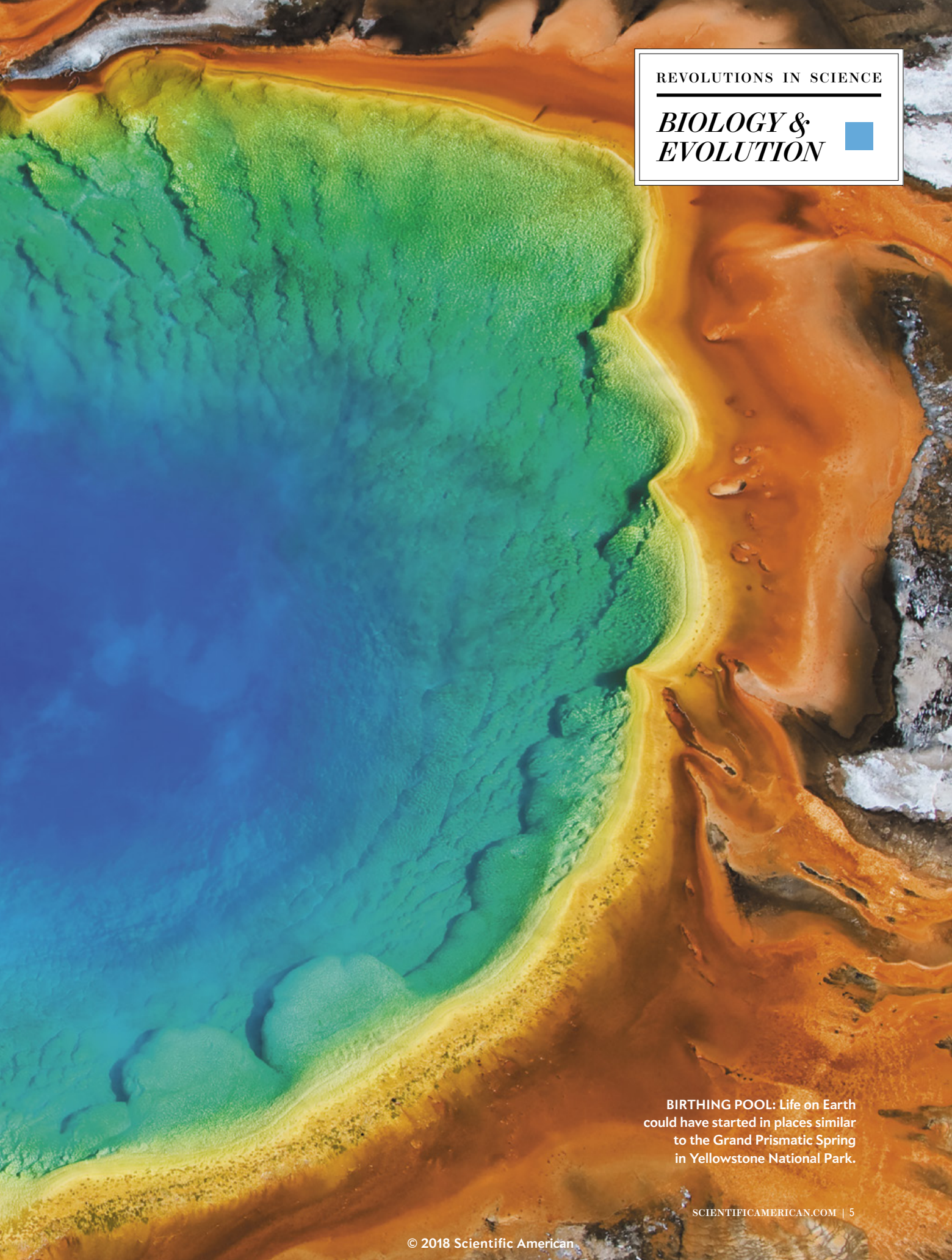


life's origins

Deep oceans were thought to hold life's origins. New evidence points instead to an active volcanic landscape

By Martin J. Van Kranendonk, David W. Deamer and Tara Djokic

© 2018 Scientific American



BIRTHING POOL: Life on Earth could have started in places similar to the Grand Prismatic Spring in Yellowstone National Park.

© 2018 Scientific American

IT'S PITCH-BLACK. WE HAVE BEEN SCRATCHING OUR WAY THROUGH DENSE underbrush in northwestern Australia, guided only by the dim light from a GPS screen. The light is too weak to reveal fallen trees that fill the dry creek bed we are following, and we keep tripping over them. We are two geologists working in a remote region of the country known as the Pilbara: Djokic up front and Van Kranendonk several steps behind. Our truck, parked somewhere on a small plateau, seems a world away. We are not sure if the GPS's batteries will hold out long enough to show us the way back. The night sky, ablaze with countless stars visible right down to the horizon, twinkles in an amazing spectacle as Jupiter dances with nearby Venus. Sadly, this spectacle provides little navigational help for two scientists fumbling their way through the Australian outback in June 2014.

Heading up the side of the creek embankment, Djokic suddenly stumbles back downhill. Has she lost her balance? To stop her from falling, Van Kranendonk reaches out to stop her and pushes her back uphill, which prompts a screech, something unintelligible, and finally a sputtered cry: "Sp- … p- … p- … pp-pider!" Djokic has not stumbled at all. She is in flight mode, in fear for her life as she tries to swat away the thick spider web enveloping her. Spiders have a deservedly bad reputation in Australia. In the dark, it is not a good idea to assume that you have found the odd benign species.

The reason we are feeling our way around the Pilbara at night is because we had spent the day enthralled by a new discovery Djokic had made in 3.48-billion-year-old sedimentary rocks called the Dresser Formation. Some of the rocks are wrinkled orange and white layers, called geyserite, which were created by a volcanic geyser on Earth's surface. They revealed bubbles formed when gas was trapped in a sticky film, most likely produced by a thin layer of bacterialike microorganisms. The surface rocks and indications of biofilms support a new idea about one of the oldest mysteries on the planet: how and where life got started. The evidence pointed to volcanic hot springs and pools, on land, about 3.5 billion years ago.

This is a far different picture of life's origins from the one scientists have been sketching since 1977. That was the year the research submarine *Alvin* discovered hydrothermal vents at the bottom of the Pacific Ocean pumping out minerals containing iron and sulfur and gases such as methane and hydrogen sulfide, surrounded by primitive bacteria and large worms. It was a thriving ecosystem. Biologists have since theorized that such vents, protected from the cataclysms wracking Earth's surface about four billion years ago, could have provided the energy, nutrients and a safe haven for life to begin. But the theory has problems. The big one is that the ocean has a lot of water, and in it the needed molecules might spread out too quickly to interact and form cell membranes and primitive metabolisms.

Now we and others believe land pools that repeatedly dry out and then get wet again could be much better places. The pools have heat to catalyze reactions, dry spells in which complex molecules called polymers can be formed from simpler units, wet spells that float these polymers around, and further drying periods that maroon them in tiny cavities where they can interact and even become concentrated in compartments of fatty acids—the prototypes of cell membranes.

What Djokic found was strong geologic evidence that the Dresser, now a dry, hot and barren outback environment, had once been like the steaming pools and erupting geysers of Yellowstone National Park in the U.S., an active geothermal field. And everywhere in the Dresser there are fossilized signs of life intimately associated with the old hot spring system. Although the Dresser was not the actual site where the most primitive life

IN BRIEF

To get started, life on Earth needed energy to create complex molecules and ways to bring these molecules together.

A system of volcanic pools and hot springs on land has the needed ingredients for life and wet-dry cycles for interaction and natural selection.

A land-based volcanic origins theory, in contrast to an ocean-focused one, guides us to different places in the solar system to search for life there.

PRECEDING PAGES: WERNER VAN STEEN *Getty Images*

© 2018 Scientific American

began half a billion years earlier, it was showing us that hydrothermal environments on land were present very early in Earth's history. Charles Darwin had suggested, back in 1871, that microbial life originated in "some warm little pond." A number of scientists from different fields now think that the author of *On the Origin of Species* had intuitively hit on something important. And the implications of these ideas stretch beyond our own planet: in our search for alien life elsewhere in the solar system, a land-based theory about origins would guide us to different places and planets than would an ocean-based theory.

FROM RUSSIA WITH LIFE

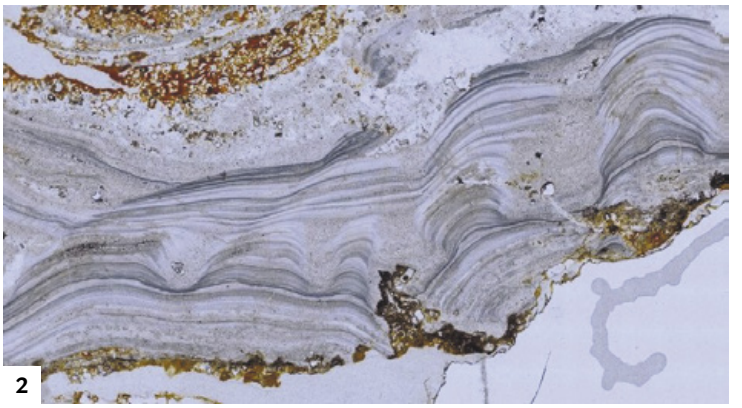
TEN YEARS BEFORE Djokic's run-in with the spider web, another of us (Deamer) had shown that volcanic pools could foster the assembly of compartments made of membranes, essential boundaries of all cellular life. Deamer led a group of scientists to Mutnovsky, an active volcano in the Kamchatka peninsula of far eastern Russia. The group was exploring a prebiotic analogue site, a region that can give researchers a sense of what the planet was like four billion years ago, before life began. Deamer's idea was that simple molecular building blocks might join into longer information-carrying polymers like nucleic acids—needed for primitive life to grow and replicate—when exposed to the wet-dry cycles characteristic of land-based hot springs. Other key polymers, peptides, might form from amino acids under the same conditions. Crucially, still other building blocks called lipids might assemble into microscopic compartments to house and protect the information-carrying polymers. Life would need all the compounds to get started, and Mutnovsky had an abundance of hot springs and geysers in which the idea could be tested.

Deamer had brought a bottle of white powder containing raw material that was likely available on the prebiotic Earth, including four amino acids and four chemical bases that compose naturally occurring nucleic acids, as well as phosphate, glycerol and a lipid. He poured this mixture into the center of a small, boiling spring. Within minutes a white, frothy foam emerged around the spring's edges. The foam was composed of countless tiny vesicles, each containing compounds that were present in the original soup.

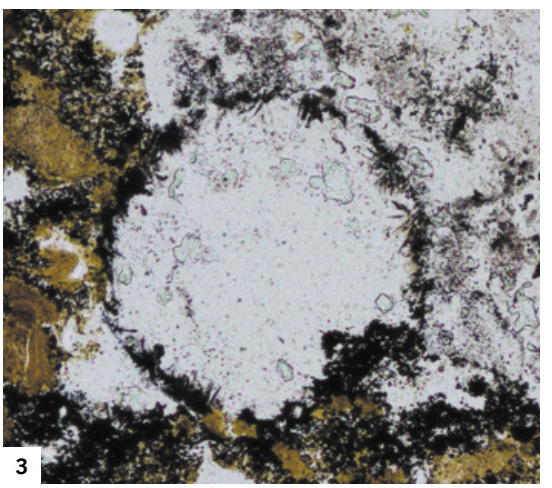
If the compartments dried out around the edges of the puddle, could their contents, already in close proximity, join together as polymers? Could this be a stepping-stone to the first life? Back in his laboratory, Deamer and his colleagues tested the idea by mixing simple nucleic acids called nucleotides with lipids. The mixture was put through cycles of wetting and drying under the acidic conditions and high temperatures found in the Kamchatka pool. The result: longer polymers ranging from 10 to more than 100 nucleotides in length. Later studies using x-ray diffraction demonstrated the polymers resembled ribonucleic acid, or RNA. Furthermore, these polymers were encapsulated by the lipids to form vast numbers of microscopic compartments called protocells. Though not alive, they were clearly an important step toward life.

Deamer used just a few wet-dry cycles in his experiments and got relatively simple molecules. A colleague of his at the University of California, Santa Cruz, computer scientist Bruce Damer, suspected that many more cycles might add another key feature: the survival of the fittest. Each drying cycle, Damer figured, would cause lipid membranes of the vesicles to open, allowing polymers and nutrients to mix. On rewetting, the lipid membranes would reencapsulate different mixtures of polymers, each mixture repre-

1



2



3

LIFE ON THE ROCKS: Orange rocks in Australia's Pilbara region are called geyserite, composed of minerals splashing from geysers in hot springs (*1*). The rocks show signature dark bands rich in titanium and light bands composed largely of potassium in a microscopic view (one centimeter in width) (*2*). Minuscule bubbles preserved in this 3.5-billion-year-old geyserite were formed in sticky biofilms, the products of biological organisms (*3*).

COURTESY OF MARTIN J. VAN KRANENDONK

© 2018 Scientific American

senting a kind of natural experiment. More complex protocells would have better chances of survival because their greater variety of molecular mixtures might stabilize the protocells in various conditions—one set of molecules helping in one set of surroundings, another helping in a different set. These intact protocells would then survive to pass on these polymer sets to the next generation, climbing an evolutionary ladder. Damer realized that this model resembled a kind of chemical computer "booting up" the functions of life, starting with random "programs" written in the form of polymers.

In 2015 Damer added a third phase to the two-part cycle: an intermediate stage between wet and dry. The idea occurred during a field trip with the coauthors to the Dresser Formation in search of stromatolites, which are the fossilized layers of bacterial mats and some of the earliest evidence of life on Earth. Damer was walking through the desert near a granite outcrop known as Gallery Hill that is covered with Aboriginal rock carvings known as petroglyphs. On the way, he noticed brown, dried-up microbial mats in small depressions in the outcrops. Out of curiosity, Damer poured water on the mats, and they sprung back to life, becoming green and gel-like. He realized that if wet-dry cycles in an origin pool also included a moist phase, in which surviving protocells crowd together into a similar gel, polymers and nutrient molecules could mix and exchange across the barriers of lipid membranes. This community of cooperating protocells would have even more opportunities to find the best molecules for survival. Forty years earlier, in fact, scientists George Fox and the late Carl Woese proposed the term "progenote" for such a communal primordial phase of life; Fox told Damer this matched his protocell gel.

POOLS OF INNOVATION

THE BUBBLES AND MINERAL composition that Djokic found in the Dresser Formation made it a likely spot for the three-part cycle to occur, and we published the evidence in May 2017 in *Nature Communications*. After we realized that the Dresser had been filled with surface hot springs in a geothermal system, it became clear that it also had contained many of the key ingredients and organizational structures required for the origin of life. It had a source of energy in the form of circulating hydrothermal fluids, rich in hydrogen, heated by magma from below. The rocks contained abundant amounts of the element boron, a crucial ingredient in the synthesis of ribose necessary for nucleic acids such as RNA. The Dresser also has phosphate minerals that dissolve out of the underlying rocks and join circulating acidic geothermal fluids. Phosphate is an important component of nucleic acids, but it is also used by all life in the form of ATP (adenosine triphosphate, the molecule that supplies energy

Genesis Landscape
Hot springs, pools and geysers can kick-start chemical systems necessary for life on Earth to begin, according to one theory. The conditions set in motion seven steps, beginning with chemical synthesis, moving through cycles of increasing complexity and ending in colonization of new territory.
1 Synthesis
Many of life's basic building blocks, such as amino acids, are formed in space and fall to Earth.
2 Accumulation
In-falling organic compounds, along with others generated within hot springs on a volcanic landscape, accumulate in hydrothermal pools.
The authors are indebted to Bruce Damer of the University of California, Santa Cruz, whose insights helped to create this diagram.

© 2018 Scientific American

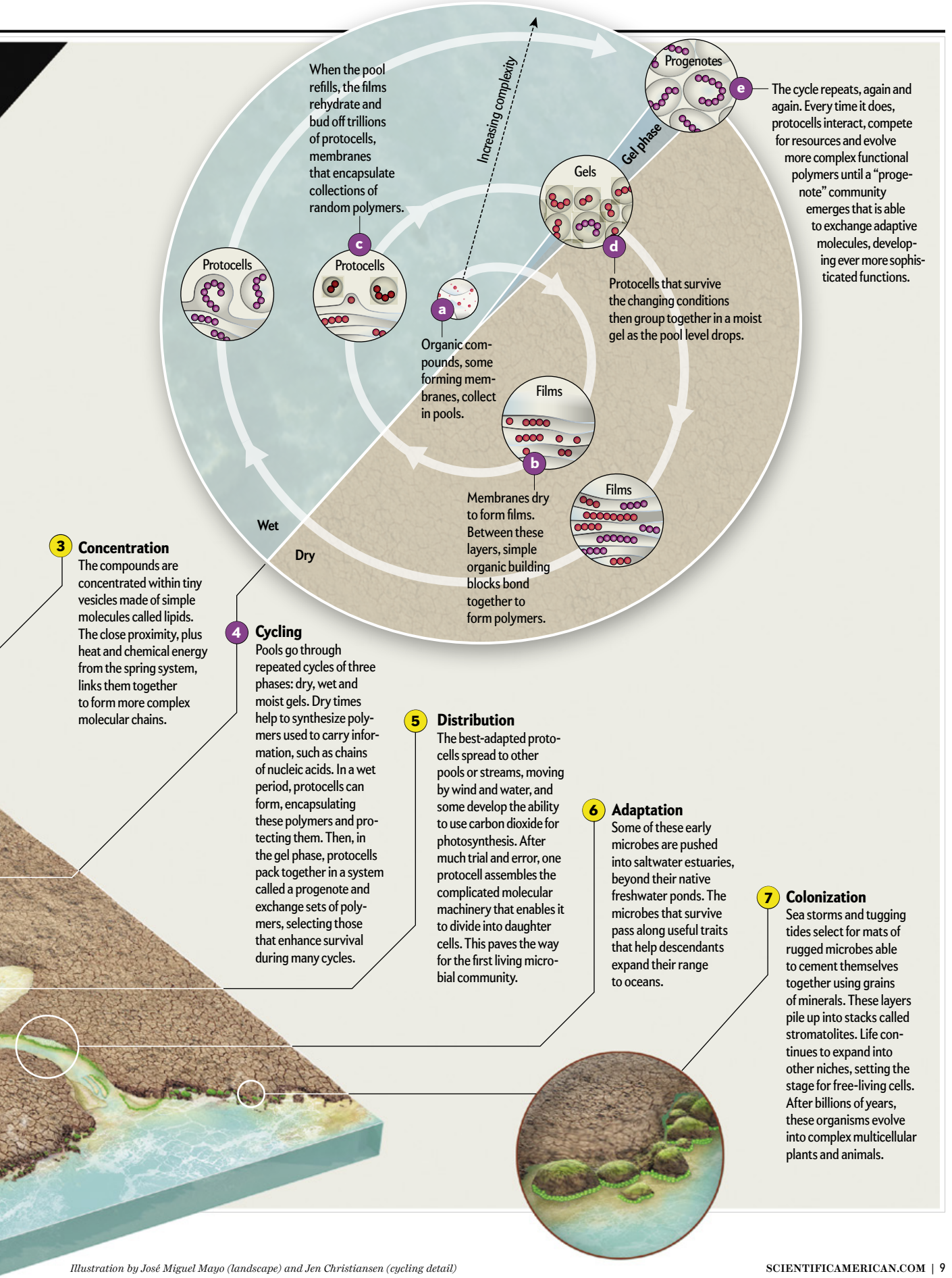

When the pool refills, the films rehydrate and bud off trillions of protocells, membranes that encapsulate collections of random polymers.
Increasing complexity
Gel phase
Progenotes
e
The cycle repeats, again and again. Every time it does, protocells interact, compete for resources and evolve more complex functional polymers until a "progenote" community emerges that is able to exchange adaptive molecules, developing ever more sophisticated functions.
Gels
d
Protocells that survive the changing conditions then group together in a moist gel as the pool level drops.
c
Protocells
Protocells
a
Organic compounds, some forming membranes, collect in pools.
Films
b
Membranes dry to form films. Between these layers, simple organic building blocks bond together to form polymers.
Films
Wet
Dry
3 Concentration
The compounds are concentrated within tiny vesicles made of simple molecules called lipids. The close proximity, plus heat and chemical energy from the spring system, links them together to form more complex molecular chains.
4 Cycling
Pools go through repeated cycles of three phases: dry, wet and moist gels. Dry times help to synthesize polymers used to carry information, such as chains of nucleic acids. In a wet period, protocells can form, encapsulating these polymers and protecting them. Then, in the gel phase, protocells pack together in a system called a progenote and exchange sets of polymers, selecting those that enhance survival during many cycles.
5 Distribution
The best-adapted protocells spread to other pools or streams, moving by wind and water, and some develop the ability to use carbon dioxide for photosynthesis. After much trial and error, one protocell assembles the complicated molecular machinery that enables it to divide into daughter cells. This paves the way for the first living microbial community.
6 Adaptation
Some of these early microbes are pushed into saltwater estuaries, beyond their native freshwater ponds. The microbes that survive pass along useful traits that help descendants expand their range to oceans.
7 Colonization
Sea storms and tugging tides select for mats of rugged microbes able to cement themselves together using grains of minerals. These layers pile up into stacks called stromatolites. Life continues to expand into other niches, setting the stage for free-living cells. After billions of years, these organisms evolve into complex multicellular plants and animals.

Illustration by José Miguel Mayo (landscape) and Jen Christiansen (cycling detail)

© 2018 Scientific American

Journey to a Land across Time

When I first set foot in Western Australia's Pilbara, a landscape holding 3.5-billion-year-old clues to the beginning of life, I was very disappointed. The year was 1994. I drove excitedly out of the west coast town of Port Hedland, but all I saw for the first 150 kilometers were a few withered, scraggly trees and smoky dust devils traipsing across the burnt, flat plain. I felt desolated. What had I gotten myself into? And the heat!! I'd never experienced anything this brutal before. Or breathed air so thick with biting flies.

But as we continued to head south on the highway to Marble Bar—the hottest town in Australia—some low, broad hills started to rise from the horizon. We started to cross sandy creeks and rivers, including the mighty Shaw, whose banks were garnished with lush-looking coolabah trees, with their distinctive, bright-white trunks.

As we continued down a dirt track into the hills, the burnt plains gave way to grass-covered hummocks. This grass is called spinifex, an amazing but devilish creation. It grows as bushes up to one meter in diameter, with round, fine blades that taper into needle-sharp tips made almost of pure silica. The tips will penetrate through just about any piece of fabric. My supervisor whipped out thick gaiters to protect his legs. But he had failed to inform me of the hazard. Without any gaiters, I was a walking porcupine within minutes—my skin skewered with multiple silica needle tips that broke off and remained in my flesh for months.

The land, ultimately, proved worth the discomfort. Here I was walking over some of Earth's oldest, best-preserved rocks that contain evidence of life from almost the very beginnings of time on our planet. As I looked at some wrinkly structures that lay above the ripples of ancient sediment, I realized I was looking at remnants of our great-, great-, great-grandparents—the precursors to all complex life on Earth!

This area had changed much from when it was first formed 3.5 billion years ago. Back then it would have been a black volcanic land, with no color from vegetation. Over the hills I might have glimpsed a green, iron-rich sea underneath an orange sky heavy with carbon dioxide and devoid of oxygen. Nearby in the landscape I'd come across fields of hot springs, and here I'd start to see some color. There would be stretches of white and yellow and red around bubbling mud pools and splashing geysers, the colors of sulfur, clay and iron. And in some pools and channels, perhaps there would have been strands of beige, red and purple: colonies of heat- and chemical-loving microbes. There might even have been some green from very early photosynthesizing organisms.

If I were able to ride a time machine forward a billion years, I'd see the Pilbara become buried under kilometers of volcanic lavas and sediments; I'd see the landmass move across the face of the globe and run into other pieces of crust, the collisions forming mountain belts. At about 2.5 billion years ago, I'd see the oceans fill with life, the shallow coastal areas occupied by huge reefs made of primitive microbes called cyanobacteria that stack in piles of mats called stromatolites. The sky would turn blue as the photosynthesizing cyanobacteria sucked in carbon and pumped out oxygen into the atmosphere. Almost another two billion years later the world would turn cold and become covered in a global ice sheet, wiping out almost every living thing. When it melted away, oxygen levels rose again. Life really got going. Animals slowly colonized the land, as did new types of plants. The greening of our planet began in earnest, and a wide variety of organisms appeared—including, unfortunately for me, spinifex.

—M.J.V.K.

CRADLE OF LIFE? Australia's Pilbara region, now dry, once held hot springs and geysers.

within cells). In addition, there were high concentrations of zinc and manganese, components of many enzymes in the cytoplasm of cells from all known branches of life, found in hydrothermal vents and in evaporative volcanic lake deposits. Finally, the Dresser also had clays, which can function as catalysts for creating complex organic molecules because of the electrically charged layers of mineral surfaces they contain.

Perhaps the most exciting thing about the Dresser as an origin analogue site is its amazing variety because in this field of science, variety is very much the spice of life. The Dresser is dry and rocky now, but in their youth, geothermal hot spring fields such as this one contain many hundreds of pools, each with a slightly different pH, temperature, dissolved ions and other chemical variations. Chemical complexity is rich in such fields because they contain three highly reactive interfaces—between water and rock, water and air, and rock and air. The fields also have different temperatures at different spots. Multiply all of this together: the wetting-drying cycles happening multiple

THEO ALLOFS *Getty Images*

© 2018 Scientific American

times each day (think Old Faithful in Yellowstone), variable pool chemistries, highly reactive interfaces, the ability of pools to exchange compounds as geysers splash their contents back and forth, and an interconnected, fluid-filled, subterranean fracture network. When you do the math, it looks as if a terrestrial geothermal field of 100 springs can generate a million or more new combinations of conditions every year!

Each warm pond becomes an "innovation pool," a test bed in which adaptive combinations of molecules rapidly emerge and find ways to grow and reproduce or in which maladaptive combinations fall by the wayside, unable to keep up. It is likely that immense numbers of combinations might have been required to assemble the first primitive version of life, in which case the process would take hundreds of millions of years. But the numbers of combinations in terrestrial geothermal fields suggest that life could have originated and begun to evolve in as little as 10 million years, with the first stages beginning as soon as there was a stable crust peppered with volcanic landmasses amid the oceans, just more than four billion years ago.

VENTING DISAGREEMENT

NOT EVERYONE AGREES that surface hot springs are the most likely sites of life's beginnings. The deep-sea vent hypothesis is still alive and kicking. At NASA's Jet Propulsion Laboratory, biochemist Mike Russell has developed *Alvin*'s original discovery of hydrothermal vents into an alternative, elegant—but as yet unproven—model. In his scheme, mineral membranes that form minuscule pores within vent rocks initially separate alkaline water from more acidic ocean water. This produces a gradient of several pH units, similar to the difference between a solution of household ammonia and a glass of orange juice. The gradient is a form of energy that can be tapped; modern bacterial cells do exactly this to generate the ATP they need. There is another source of energy in the vents in the mixture of dissolved gases such as hydrogen and carbon dioxide. Russell and his colleagues have proposed that when carbon dioxide in ancient seawater mixed with hydrogen coming from the vents, the transfer of electrons from hydrogen to carbon dioxide could synthesize more complex organic compounds. In their view, the mineral compartments resemble cells, and the energy of pH gradients and hydrogen could ultimately evolve into a primitive metabolism required by the earliest forms of life.

The hot spring field and deep-sea vent hypotheses have some far-flung implications. Beyond guiding further explorations of life's beginnings on Earth, they point to different approaches to search for life on other planets and their moons. If the deep-sea vent origins theory is correct, the icy ocean worlds of Saturn's Enceladus and Jupiter's Europa may be good places to look. On the other hand, if our model of fluctuating hot springs is right, then these worlds are unlikely to host life.

What about Mars? Although there is good evidence for shallow seas on Mars in the distant past, there are few signs of a global ocean or of tectonic spreading zones that create hydrothermal vents on Earth. If life depended on vents to begin, it was unlikely to have begun on the Red Planet. But if life on Earth originated in terrestrial hot springs, it could have also begun on Mars, which had the hot spring ingredients of widespread volcanism and water. Indeed, in 2008 the Spirit rover discovered 3.65-billion-year-old hot spring deposits in the Columbia Hills on Mars, about the same age as our Dresser hot springs, which did a great job of preserving early evidence for life on Earth.

Both the deep-sea vent and the land-based hot spring pools models have a long way to go before either can be deemed correct. The origin of life is like a jigsaw puzzle with many different pieces, and we do not know enough yet to put each one in the proper position. At the Dresser Formation, for instance, we do not understand what causes certain elements to become concentrated in different pools, how geothermal fields evolve over time, or how their different chemistries interact to synthesize or degrade organic molecules. We need to construct more sophisticated experiments of prebiotic chemistry in a series of warm little pools, studying how complex organic molecules form and how they interact and combine when encapsulated within membranes.

Both on land and in the sea, chemical and physical laws have provided a very useful frame around this particular puzzle, and the geologic and chemical discoveries described here fill in different areas. But before we can see a clear picture of the origin of life, many more pieces need to be put in place. What is exciting, however, is that now we can see a path forward to the solution. SA

Martin J. Van Kranendonk is director of the Australian Center for Astrobiology in the School of Biological, Earth and Environmental Sciences at the University of New South Wales. He has conducted research for more than 30 years in extremely old rocks across the planet.

David Deamer is a faculty member in the department of biomolecular engineering at the University of California, Santa Cruz. He is author or editor of 12 books, including *The Origins of Life* (2010), co-edited with Jack W. Szostak, and *First Life* (2011), published by the University of California Press.

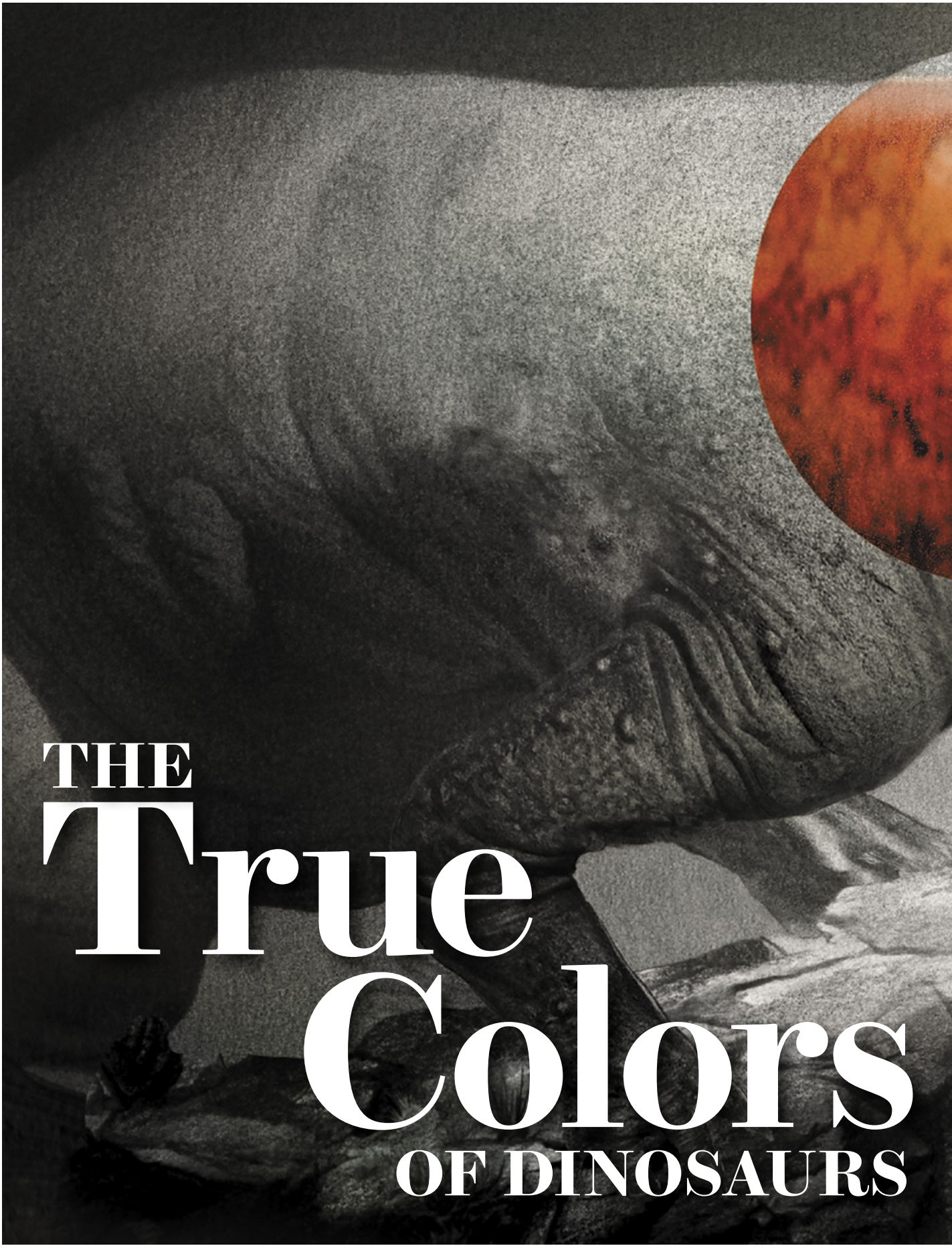
Tara Djokic is a Ph.D. candidate at the Australian Center for Astrobiology at the University of New South Wales. Her project combines geologic observations of early evidence of life in Western Australia with virtual-reality technology.

MORE TO EXPLORE

A Nonhyperthermophilic Common Ancestor to Extant Life Forms. Nicolas Galtier et al. in *Science,* Vol. 283, pages 220–221; January 8, 1999.
The Onset and Early Evolution of Life. Michael J. Russell and Allan J. Hall in *Geological Society of America Memoirs,* Vol. 198, pages 1–32; 2006.
Geological Setting of Earth's Oldest Fossils in the ca. 3.5 Ga Dresser Formation, Pilbara Craton, Western Australia. Martin J. Van Kranendonk et al. in *Precambrian Research,* Vol. 167, Nos. 1–2, pages 93–124; November 10, 2008.
First Life: Discovering the Connections between Stars, Cells, and How Life Began. David Deamer. University of California Press, 2011.
Molecular and Cellular Fossils of a Mat-like Microbial Community in Geothermal Boratic Sinters. Wriddhiman Ghosh et al. in *Geomicrobiology Journal,* Vol. 29, No. 10, pages 879–885; 2012.
Origin of First Cells at Terrestrial, Anoxic Geothermal Fields. Armen Y. Mulkidjanian et al. in *Proceedings of the National Academy of Sciences USA,* Vol. 109, No. 14, pages E821–E830; April 3, 2012.
Ester-Mediated Amide Bond Formation Driven by Wet-Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. Jay G. Forsythe in *Angewandte Chemie International Edition,* Vol. 54, No. 34, pages 9871–9875; August 17, 2015.
Hydrothermal Conditions and the Origin of Cellular Life. David W. Deamer and Christos D. Georgiou in *Astrobiology,* Vol. 15, No. 12, pages 1091–1095; December 2015.
A Field Trip to the Archaean in Search of Darwin's Warm Little Pond. Bruce Damer in *Life,* Vol. 6, No. 2, Article No. 21; June 2016.
Earliest Signs of Life on Land Preserved in ca. 3.5 Ga Hot Spring Deposits. Tara Djokic et al. in *Nature Communications,* Vol. 8, Article No. 15263; May 9, 2017.

scientificamerican.com/magazine/sa

© 2018 Scientific American



THE True Colors OF DINOSAURS

© 2018 Scientific American

PSITTACOSAURUS is one of several dinosaurs whose coloring has been deduced from fossilized pigments.

Long thought impossible, preservation of fossil pigments is allowing scientists to reconstruct extinct organisms with unprecedented accuracy—a feat that is yielding surprising insights into the lives they led

By Jakob Vinther

Illustration by Greg Ruth

© 2018 Scientific American

ON A DAY IN OCTOBER 2006, I SAT IN A DARK LABORATORY AT YALE University and zoomed into the fossilized ink of a 200-million-year-old squid relative under an electron microscope. An ocean of translucent balls, each roughly a fifth of a micron in diameter, loomed into view. To the untrained eye, they might have been unimpressive. But I was riveted. These ancient structures looked exactly like the granules of melanin pigment that color the ink of modern squid and octopuses.

Perhaps I should not have been so surprised at the resemblance. Researchers had announced the first discovery of fossil ink granules a couple of years earlier. But seeing them with my own eyes was a revelation. As I examined cephalopod specimens from various locales and time periods, I realized their ink was always the same, perfectly preserved for hundreds of millions of years.

The consistently superb preservation of the ink made me wonder whether melanin might persist in fossils of other kinds of organisms. Melanin is the same pigment found in hair, skin, feathers and eyes. It can impart red, brown, gray and black hues and create metallic sheens. If I could find melanin in other fossils, perhaps I could reconstruct the coloring of extinct animals, including dinosaurs. For decades scientists have assumed that pigments hardly ever survive the fossilization process. The few known examples all came from fossils of invertebrate creatures, not backboned ones. Thus, researchers could only guess at the colors of most long-vanished animals, using modern ones as a guide. As a result, dinosaur reconstructions varied widely: some sport the drab earth tones associated with reptiles and amphibians; others flaunt the rainbow hues of modern birds (the only dinosaurs that have survived to modern times).

But discoveries I and others have made over the past dozen years are taking out some of the guesswork. Our examinations of dozens of fossils have revealed many examples of melanin-bearing structures. By studying the shapes and organizations of these structures, we have been able to deduce the actual colors and patterns of extinct dinosaurs and other animals from deep time. These clues to the physical appearances of the creatures, in turn, have led to intriguing insights into their behaviors and habitats.

To test my hypothesis that melanin survives in other fossils and can be used to deduce the true colors of extinct animals, I wanted to find and analyze fossils with dark stains indicative of organic preservation in those anatomical regions generally known to contain melanin: the outer covering of the body and the eyes. And I needed to be able to examine the darkened areas under the electron microscope, which might require cutting a specimen down to size. Well-preserved fossils are rare, however, and museums guard them closely. Fortunately, a remarkable fossil site in my home country of Denmark called "Fur and Ølst Formation" had yielded exquisite bird fossils with feathers, which would be an ideal test case. I managed to convince the curator of vertebrate fossils at the Geological Museum in Copenhagen to cut down a typewriter-sized block of limestone containing a skull of a little bird with stains where the eyes used to be and a dark halo of feather impressions into a piece the size of a slice of bread so that it could fit into the museum's electron microscope.

I had a good idea of what to search for under the microscope. Before obtaining the fossil bird for analysis, I had read numerous scientific papers to figure out what melanin looks like in the feathers of living birds. Melanin is synthesized in specialized cells known as melanocytes by cellular components called melanosomes. Typically the melanin remains encased in the melanosomes, which measure about 0.5 to two microns long and take two forms: a sausage-shaped kind that produces a form of melanin called eumelanin, which absorbs all wavelengths of light and thus gives squid ink and raven feathers their black color, and a meatball-shaped variant that makes pheomelanin, which imparts a rusty red hue. An absence of pigments results in white plumage. Gray and brown colors, for their parts, appear to arise from combinations of eumelanin, pheomelanin and pigment absence.

I had also consulted one of the world's leading experts on bird color: Richard Prum of Yale. Because I knew from the fossil ink that eumelanin can preserve, I figured I would start by looking for that pigment in the feathers. Talking to Prum and his then Ph.D. student Vinod Saranathan, I learned that the sausage-shaped melanosomes line up in a distinctive way along the barbs and barbules that constitute a feather's branches. The melanosomes arrive there during development, when the melanocytes transfer them into specialized cells called keratinocytes

IN BRIEF

Scientists long assumed they could only guess at the colors of dinosaurs and other extinct organisms.
But recent discoveries of preserved pigments in fossils of a wide range of creatures have upended that notion.
Analyses of the pigments are allowing researchers to infer the colors of animals that vanished long ago.
The color patterns have, for their part, revealed other previously unknown aspects of the animals' lives.

© 2018 Scientific American

that give rise to feathers and hair. If the dark stains on the feather impressions evident in the Danish bird fossil came from melanin, then I should see the sausages arranged this way along the feather branches under the microscope.

With great anticipation, I zoomed in on the fossil feathers—and encountered millions of sausage-shaped structures. Unfortunately, the underground railway was less than 50 meters from the museum's basement, where the electron microscope was located; vibrations from the constant train traffic made it impossible to get a clear image. But the images were good enough to see the sausages. I immediately e-mailed them to my then Ph.D. supervisor at Yale, Derek Briggs, a pioneer in the study of extraordinarily preserved fossils. He replied with less enthusiasm than I had hoped for, noting that these structures were the same as those he and others had found in fossil feathers and mammal hair for decades and had identified as bacteria.

I still thought the sausages were melanosomes, though, and made my argument to Briggs. Not only did they have the right shape and size but their orientation in the feather structures mirrored that of black melanosomes in modern bird feathers. Furthermore, it was clear from the fossil squid ink that melanin can fossilize. Briggs began to warm to the idea, but he was not convinced until he showed the images to Prum, who confirmed that they resembled melanosomes in every aspect.

To bolster the hypothesis that melanosomes can persist in fossils of extinct birds, Briggs wanted to find another example. He rummaged through the scientific literature for a good test case and found a description of a little Cretaceous feather from Brazil that preserves distinct black and white color bands. Briggs thought that if we could show that this specimen also preserves aligned melanosomes—but only in the dark bands because white coloration stems from a lack of pigment—we would have enough evidence to make our case. We managed to get the specimen on loan and put the entire block under the electron microscope. Lo and behold, when I examined the dark bands of this 108-million-year-old feather, thousands of little melanosomes aligned along the axes of the fine feather branches came into focus. When I looked at the white bands, in contrast, I saw nothing but rock matrix—which is exactly what one should expect in the absence of pigment.

FOSSIL of a *Psittacosaurus* **preserves pigment patterns indicative of a type of camouflage called countershading.**

COURTESY OF JAKOB VINTHER AND BOB NICHOLLS

PAINT BY NUMBER

SINCE THE PUBLICATION of our melanosome discoveries in 2008, my team and several others have described melanosomes and other pigments from additional fossils. Researchers have also started investigating the chemistry of fossil melanin and substantiated our observations that melanin can survive for millions of years, almost chemically intact. Together with Caitlin Colleary, then a master's student at the University of Bristol in England, where I now work, we showed that the slight alterations evident in the fossil melanin are the result of sustained exposure to elevated pressure and heat in the ground. (A few investigators still maintain that the observed structures might be bacteria, but they are running out of options to support their claims.)

Some of our most spectacular findings have uncovered the colors of dinosaur feathers. In 2009 my Yale colleagues and I teamed up with Matthew Shawkey and Liliana D'Alba, both now at Ghent University in Belgium, and others to reconstruct the color pattern of *Anchiornis huxleyi,* a small, predatory, feathered dinosaur from China that lived around 155 million years ago. Like the Danish bird I had studied previously, the *Anchiornis* fossil had some dark stains visible to the naked eye, indicating the presence of organic material, probably melanin. But because we were aiming to reconstruct the pattern of its full plumage—a much more ambitious task than simply determining the presence or absence of melanosomes—we could not rely on these stains to tell us all we wanted to know. Instead we had to develop a way to objectively predict colors from the shapes of the melanosomes. To do this, we studied melanosomes from 12 black, 12 brown and 12 gray feathers of modern-day birds. By considering the length, width and aspect ratio of the melanosomes, as well as how much they vary in shape, we could predict feather color using a statistical method called quadratic discriminant analysis with 90 percent accuracy.

When we applied our method to the melanosomes of *Anchiornis,* the results were striking. Our statistical predictions indicated that the feathers that covered much of the creature's body were mostly gray. The long feathers on the animal's arms and legs, in contrast, were unpigmented by melanosomes and thus white, except for the melanosome-laden tips, which we predicted were black. (Modern birds often have black-tipped wing feathers. The melanin, in addition to coloring the feathers, also fortifies them against battering winds. Perhaps *Anchiornis* benefited from this strengthening property of melanin, too.) Most surprisingly, the feathers on the crown of the head contained impressions of round melanosomes—the "meatballs"—that would have given *Anchiornis* a ruddy crest. All told, this combination of colors made for a spectacularly flamboyant creature.

At around the same time we published our *Anchiornis* study, Fucheng Zhang of the Institute of Vertebrate Paleontology and Paleoanthropology in Beijing, Michael J. Benton of the University of Bristol and their colleagues reported that they had found fossil melanosomes in a range of birds and dinosaurs recovered from 130-million-year-old rocks in China. The pattern of meat-

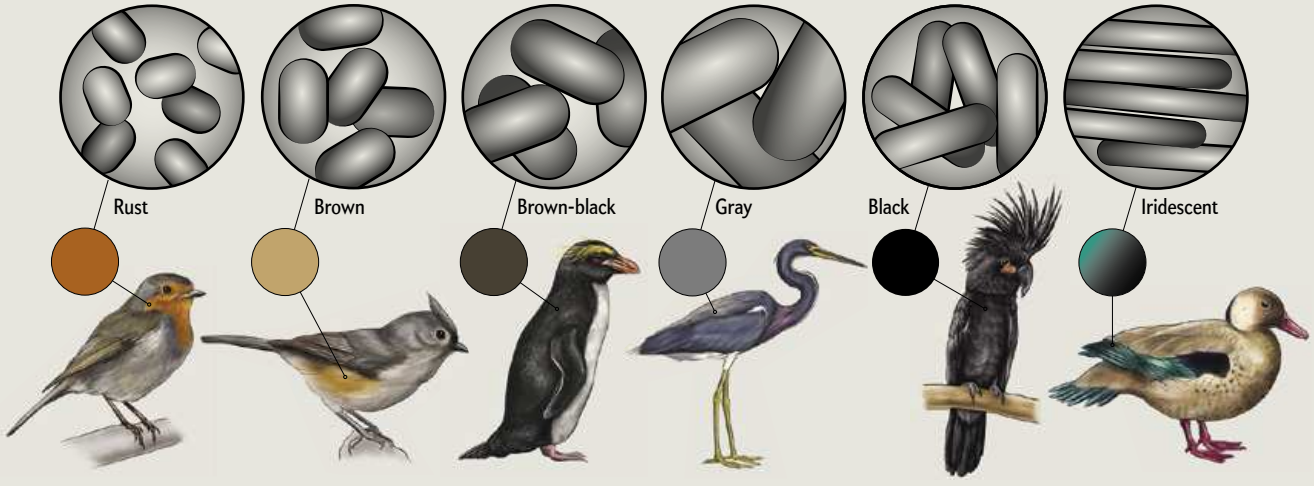
© 2018 Scientific American

In Living Color

Microscopic pigment-bearing cell structures known as melanosomes can persist in fossils for tens of millions of years. Studies of preserved pigments have allowed scientists to reconstruct the actual colors of a wide range of extinct animals, including a number of dinosaurs. These findings are not only revealing, for the first time, what these creatures really looked like, but they are also elucidating previously murky aspects of the animals' lives—from their activity cycles to the type of environment they inhabited.

Melanosome Size, Shape and Configuration Predict Color

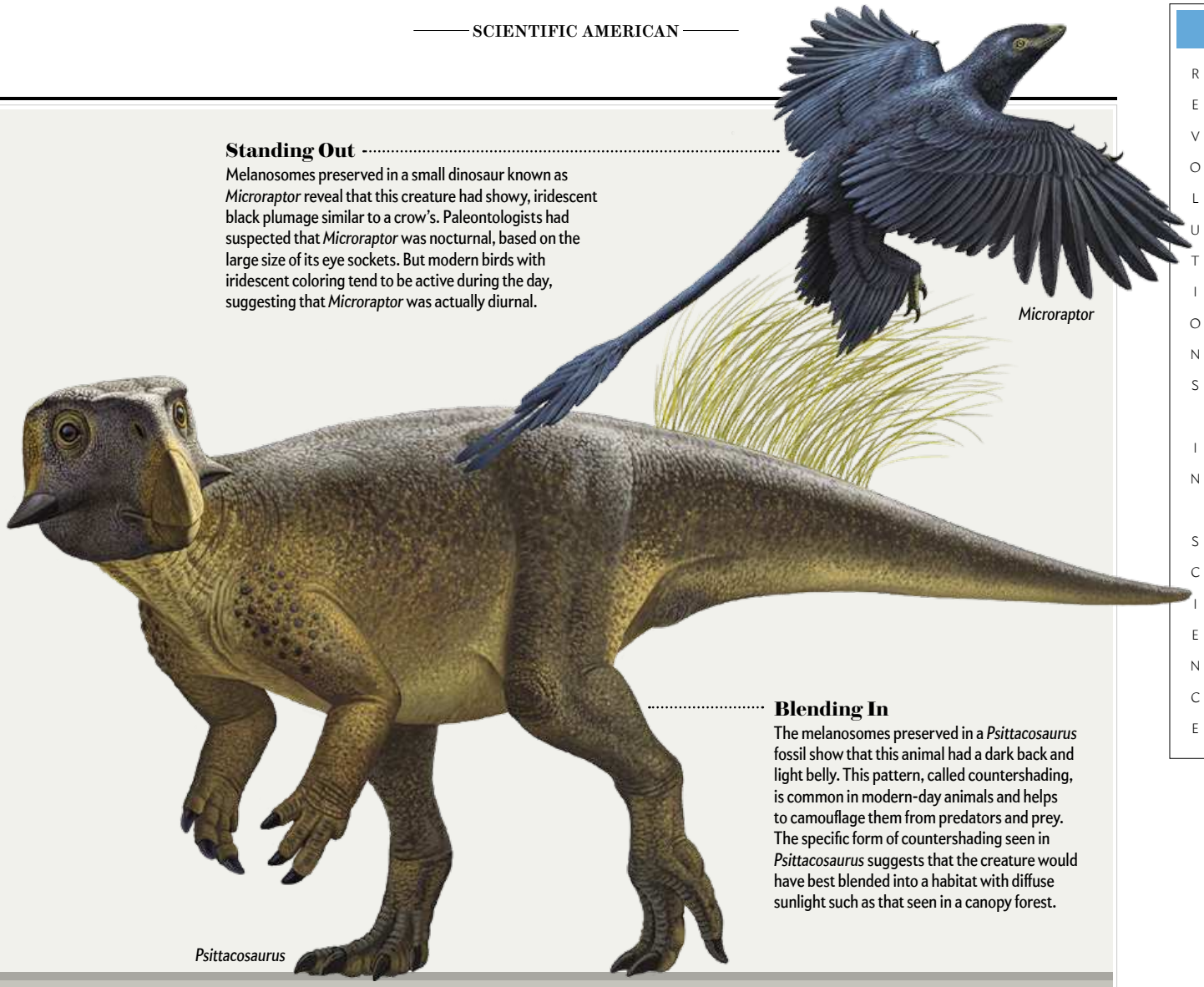
Melanosomes contain two forms of melanin: eumelanin, which gives rise to black tones, and pheomelanin, which imparts rusty red hues. Combinations of these melanins and absence of pigment create brown, gray and white colors. Iridescence, for its part, stems from the stacking of melanosomes in ways that refract light. Analyses of melanosomes from feathers of modern-day birds have yielded a database that researchers can use to predict colors and patterns of extinct animals from the size, shape and arrangement of fossil melanosomes.



© 2018 Scientific American

Standing Out

Melanosomes preserved in a small dinosaur known as *Microraptor* reveal that this creature had showy, iridescent black plumage similar to a crow's. Paleontologists had suspected that *Microraptor* was nocturnal, based on the large size of its eye sockets. But modern birds with iridescent coloring tend to be active during the day, suggesting that *Microraptor* was actually diurnal.



Microraptor

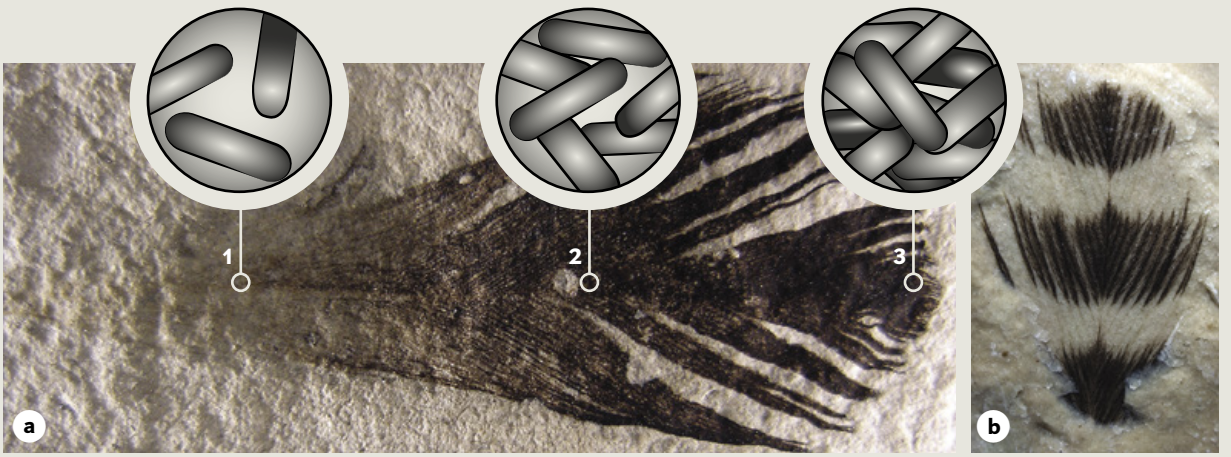
Psittacosaurus

Blending In

The melanosomes preserved in a *Psittacosaurus* fossil show that this animal had a dark back and light belly. This pattern, called countershading, is common in modern-day animals and helps to camouflage them from predators and prey. The specific form of countershading seen in *Psittacosaurus* suggests that the creature would have best blended into a habitat with diffuse sunlight such as that seen in a canopy forest.

Melanosome Density and Distribution Predict Pattern

Fossil feathers show how varying degrees of melanosome concentration can create patterns. For instance, the gradient pattern seen in a 55-million-year-old specimen from Denmark **a** arises from the combination of low melanosome concentrations that yield pale colors (*1*), intermediate concentrations that produce midrange tones (*2*) and high concentrations that form intense tones (*3*). In a 108-million-year-old fossil feather from Brazil **b**, dark and light stripes stem from melanosome-rich and melanosome-free areas, respectively.



COURTESY OF JAKOB VINTHER *(feather photographs)*

Illustrations by Raúl Martín (dinosaurs) and Jillian Ditner (birds and melanosomes)

© 2018 Scientific American

ball melanosomes in one fuzz-covered dinosaur, *Sinosauropteryx,* implied that it had sported a reddish coat and a tiger-striped tail, making it the first known ginger dinosaur.

Since those early days our feather data set has grown to comprise hundreds of samples, including ones that allow us to accurately predict iridescence, the metallic sheen seen in the plumage of hummingbirds and peacocks, among other birds. Melanosomes responsible for this effect tend to be longer than typical melanosomes, and they may even be hollow or flattened. The iridescence arises from the packing of the melanosomes within the feather. Certain configurations of melanosomes refract light in ways that create different colors, depending on the angle at which the animal is viewed or illuminated.

Amazingly, in 2009 we found evidence of iridescence in a 49-million-year-old fossil feather from Messel, Germany. The fossil, kept at the Senckenberg Museum in Frankfurt, preserves the original arrangement of melanosomes that generated the iridescence. They were packed into a dense, smooth layer found in the finest branches of the feather fossil, the barbules. There the melanosomes occurred strictly on the farthest edge of the feather and on the top surface, the only part that was not obscured by other, overlapping feathers. We deduced that the tips were iridescent because that arrangement of melanosomes is known to produce what is called thin-film interference, the kind that occurs when gasoline floats on water and creates a vivid rainbow of colors.

THE PATTERNS WERE SUBTLE, WITH FINE VEINING, DOTS AND STRIPES.

It was not long before we discovered evidence of iridescence in an actual dinosaur—a crow-size creature from China with wings on all four limbs. Dubbed *Microraptor,* it was a primitive cousin to *Jurassic Park*'s *Velociraptor.* The movie depicted *Velociraptor* with scaly skin, but scientists now know that both these dinosaurs were, in fact, covered in feathers. In *Microraptor,* the feathers preserve long, sausage-shaped melanosomes arranged to bend light in eye-catching ways. Its plumage thus would have been black, with the same shiny sheen as a crow's. *Microraptor* is not the only extinct creature now known to have had that rainbow shimmer. Jennifer Peteya of the University of Akron and Ghent's Shawkey recently described shimmering iridescence in an enantiornithine bird, called *Bohaiornis,* and a Jurassic theropod with a big, fan-shaped tail, named *Caihong.*

MORE THAN SKIN DEEP

BEYOND ALLOWING PALEONTOLOGISTS and artists to reconstruct extinct organisms more accurately, fossil pigments are revealing previously unknown facets of the daily lives of both dinosaurs and other long-gone creatures. For instance, experts had presumed that *Microraptor* was nocturnal, based on the large size of its eye sockets. But our discovery that it possessed iridescent plumage suggests otherwise because in modern birds such coloration is typically found in species that are active in the daytime. The bold coloring of *Anchiornis,* for its part, probably helped to attract mates or served as some other kind of display, as occurs in flashily dressed modern birds. Thus, color patterns may provide a way to test behavioral hypotheses about a species using a different line of evidence than usual.

Fossil pigments in one species can also illuminate aspects of the other species with which it interacted. Among insects, most color patterns evolved not to help the creatures attract mates but rather as a tactic to avoid getting eaten. Their pigments can thus provide clues to their predators. Fossils of insects called lacewings offer a fascinating example. Between 170 million and 150 million years ago certain distinctive color patterns made their evolutionary debut in insects. Perhaps the most dramatic pattern to emerge during this time was the eyespot, a marking that resembles the eye of a different kind of animal and serves to startle predators approaching their prey at speed from a distance. Lacewings are one of the first creatures known to have had eyespots. What kind of predator were they defending against? Most color patterns of modern insects have evolved as a defense against birds, which are their main predators nowadays. But the lacewings' eyespots predate the origin of birds as we know them. Their predators were instead most likely a small group of dinosaurs called the paravians, which are known to have lived at the same time as these lacewings and are thought to have given rise to birds. Although the fossil record of paravians themselves has been unable to unequivocally pinpoint when flight evolved in this group, the appearance of these eyespots in the lacewings hints that some paravian dinosaurs had taken wing by this point and were exerting birdlike predation pressure on the insects.

Other fossil melanosome discoveries have allowed my collaborators and me to reverse engineer the environment in which extinct organisms lived. Our first foray into this realm of investigation began with a particularly splendid fossil of a small, plant-eating dinosaur called *Psittacosaurus,* a relative of *Triceratops.* These skeletons are quite common in northeastern China and are often very complete. This specimen stood out even in that good company, however. A thin film drapes its body—the remains of the skin, including delicate scales. And its tail displays long, filamentous bristles that may be precursors to feathers. Previous discoveries of dinosaur feathers have all come from the mostly carnivorous theropod group of dinosaurs. The bristles on *Psittacosaurus,* a distantly related member of the plant-eating ceratopsian group, hint that plumage might have been far more widespread among the dinosaurs than previously thought.

When I first encountered the specimen in 2009, a year after we had announced the discovery of melanosomes in fossil birds, I saw right away that it preserved evidence of beautiful color patterns all over the body. The patterns were subtle, with fine veining, dots and stripes. And I could see that the animal had a dark back that gave way to a pale belly. That kind of dark-to-light color gradient from back to belly counteracts the light-to-dark gradient created by illumination from the sun. This pattern, known as countershading, is common among modern animals ranging from dolphins to deer, helping both predators and prey blend in with their surroundings and thereby elude detection.

I eventually showed the *Psittacosaurus* pattern to Innes Cuthill, who is part of a group that studies camouflage at the University of Bristol. It was then that we realized that we had the opportunity not only to study countershading in a dinosaur but also to deduce from the fossil alone what kind of environ-

© 2018 Scientific American

ment the creature lived in. To reconstruct an animal's habitat, scientists usually gather clues from fossils of other animals and plants found nearby. This kind of approach is problematic, however, because oftentimes the site where a fossil is discovered is not where the organism lived. The Chinese psittacosaur, for example, was recovered from sediments of an ancient lake. The creature was clearly not aquatic, so its remains must have been transported to the lake from the surrounding terrestrial environment, perhaps by moving water. Our study might be able to provide clues about that setting—specifically, the light conditions under which this dinosaur evolved its camouflage.

Cuthill and his collaborators had recently studied countershading in modern ungulates; the group that includes horses, antelope, camels, pigs and rhinoceroses. Although countershading by definition involves darker coloration on the back and lighter coloration on the underside (except for some animals, such as caterpillars, that live their lives upside down), the intensity of those shades and the nature of the transition from dark to pale differ from species to species. Cuthill's team wanted to investigate how well that variation correlates to variation in the lighting conditions found in different environments. Because sunlight varies depending on the latitude at which an animal lives, as well as the density of vegetation in its habitat, the researchers had theorized that ungulate countershading, too, should differ according to latitude and habitat. Their findings bore out that notion. Broadly speaking, if an animal lives in open habitats, the direct sunlight will create a shadow high on the body, with a very sharp transition to the illuminated areas. These animals usually exhibit a countershading that matches this pattern, with dark backs that almost immediately give way to pale bellies and little intermediate coloration in between. Pronghorn antelope offer a great example of this kind of countershading. In closed habitats, in contrast, the diffuse light that filters down through the vegetation scatters in all angles, producing a shadow that hangs farther down the body and transitions to the illuminated area gradually. White- and black-tailed deer, common in North American forestlands, exhibit this pattern.

We knew from our visual inspection of the *Psittacosaurus* fossil that it had countershading of some sort. We therefore carefully projected the pigment pattern onto an accurate, life-size model of the dinosaur, which we accomplished by enlisting the help of British paleoartist Bob Nicholls. Through this work we determined that the transition from dark to light occurred low on the belly and tail in *Psittacosaurus*.

To test the function of the dinosaur's color pattern, we painted a second copy of the full-scale model gray. We then photographed this model in a range of daylight conditions, from gloriously sunny to oppressively cloudy, as well as in open land and underneath conifer trees to capture the shadows cast on it. Next we inverted the dark and light shades in the photographs, effectively creating the ideal countershading patterns for concealing the animal in each of the lighting conditions. Comparing our reconstruction of the actual countershading pattern of the *Psittacosaurus* with the idealized countershading patterns, we determined that the animal's coloring would have best camouflaged it in a habitat with diffuse light, such as that seen in a canopy forest.

Subsequent studies on countershading have found that the tiger-striped *Sinosauropteryx* was adapted to live in open environments with bright sunshine from above and that it carried a striped tail and a bandit mask of pigmented feathers over its eyes. *Psittacosaurus* and *Sinosauropteryx* are found in the same fossil beds, but their countershading tells us that they came from different environments.

A nearly six-meter-long ankylosaur, *Borealopelta,* was found in the approximately 112-million-year-old marine oil sand beds in northern Alberta, Canada. It was colored with reddish-brown pheomelanin and was countershaded. Living land animals of such size are not countershaded, because there are no predators big enough to threaten them. In other words, for such a big creature to maintain its countershading from generation to generation, the Cretaceous predators must have been vicious enough to have threatened it. Perhaps that doesn't come as a big surprise considering the spine-chilling Herculean dimensions of the theropod predators back then. But now we have rock-solid proof in fossil color pattern of their gruesomeness.

A VIVID FUTURE

SCIENTISTS STILL HAVE MUCH TO LEARN about paleocolor. Our ability to see broad categories of color in fossils—those that stem from the shape and arrangement of melanosomes—is already a massive leap forward from what we knew about ancient hues less than 10 years ago. But there are other pigments to look for in fossils, including carotenoids, which produce bright reds and yellows, and porphyrins, which produce such hues as green, red and blue. These pigments have turned up in the fossil record on occasion. Researchers have identified carotenoid pigments derived from fossil bacteria dating back several billion years; porphyrins are preserved in a blood-engorged mosquito from 46 million years ago and in the eggs of a 66-million-year-old dinosaur known as an oviraptorosaur. Pigments not known from modern organisms have come to light, too, including some from fossil sea lilies and algae dating to between 300 million and 150 million years ago.

We will probably encounter limitations to the detail with which we can reconstruct paleocolors; over millions of years some information is bound to be lost forever. In addition, because exceptional fossils with organic preservation are rare and precious, we must restrict destructive chemical sampling of them. As techniques advance, however, the new discoveries they afford will undoubtedly change our understanding of the past faster than ever before. Each one will bring us that much closer to seeing dinosaurs and other prehistoric creatures as they really were, in full Technicolor glory. SA

Jakob Vinther thought he was going to become a botanist until he found his first fossils at age 11 while attending summer camp in his home country of Denmark. Today he is a senior lecturer in the schools of biological sciences and earth sciences at the University of Bristol in England. His research focuses on pigments and other molecules preserved in the fossil record.

MORE TO EXPLORE

The Colour of Fossil Feathers. Jakob Vinther et al. in *Biology Letters,* Vol. 4, No. 5, pages 522–525; October 23, 2008.

3D Camouflage in an Ornithischian Dinosaur. Jakob Vinther et al. in *Current Biology,* Vol. 26, No. 18, pages 2456–2462; September 26, 2016.

scientificamerican.com/magazine/sa

© 2018 Scientific American

THE NEW ORIGINS OF TECHN

ARCHAEOLOGISTS at a site in northwestern Kenya called Lomekwi 3 have unearthed the oldest stone tools in the world.

Photographs by Andrew Renneisen

© 2018 Scientific American

OLOGY

ANCIENT STONE TOOLS FROM KENYA SHATTER THE CLASSIC STORY OF WHEN AND HOW HUMANS BECAME INNOVATORS

By Kate Wong

© 2018 Scientific American

The desert badlands on the northwestern shores of Kenya's Lake Turkana offer little to the people who live there. Drinking water is elusive, and most of the wild animal populations have declined to near oblivion. The Turkana scrape by as pastoralists, herding goats, sheep, cattle, donkeys and the occasional camel in the hot, arid countryside. It is a hard life. But millions of years ago the area brimmed with freshwater, plants and animals. It must have been paradise for the human ancestors who settled here.

2

Sonia Harmand has come to this region to study the legacy these ancestors left in stone. Harmand is an archaeologist at Stony Brook University. She has an intense gaze and a commanding presence. On a hazy July morning Harmand sits at a small, wood folding table, scrutinizing a piece of rock. It is brownish-gray, about the size of her pinkie fingernail, and utterly unremarkable to the untrained eye. But it is exactly what she has been looking for.

Nearby 15 workers from Kenya, France, the U.S. and England are digging their way into the side of a low hill. They tap hammers against chisels to chip away at the buff-colored sediments, searching for any bits of rock that could signal ancient human activity. At the top of the hill, the workers' water bottles hang like Christmas ornaments on the thorny branches of an acacia tree; the early breeze will keep their contents cool a little longer before the heat of the day sets in. By afternoon the air temperature will top 100 degrees Fahrenheit, and the excavation floor, windless and sun-cooked, will live up to its nickname: the Oven.

In 2015 Harmand and her husband, Jason Lewis, a paleoanthropologist at Stony Brook, announced that their team had discovered 3.3-million-year-old stone tools at this site, which is called Lomekwi 3. They were the oldest stone tools ever found by far—so old that they challenged a cherished theory of human evolution. The scientists want to learn who made the tools and why. But they also have a more immediate task: unearthing more evidence that the tools are, in fact, as old as they appear.

The fragment in Harmand's hand is the first evidence of ancient stone-tool production the researchers have recovered since they got here. It is a piece of debris produced by knapping—the act of striking one rock against another to produce a sharp-edged flake. Small and light, the fragment implies that the site has not been disturbed by flowing water in the millions of years since. That fact, in turn, supports the argument that the Lomekwi 3 tools come from this ancient sedimentary layer and not a younger one. Now that the excavators have hit the artifact-bearing level of the site, they must proceed with care. "*Pole pole,*" Harmand instructs them in Swahili. Slowly, slowly.

Paleoanthropologists have long viewed stone-tool production as one of the defining characteristics of the *Homo* genus and the key to our evolutionary success. Other creatures use tools, but only humans shape hard materials such as rock to suit their purposes. Moreover, humans alone build on prior innovations, ratcheting up their utility—and complexity—over time. "We seem to be the only lineage that has gone fully technological," says Michael Haslam of the University of Oxford. "It isn't even a crutch. It's like an addition to our bodies."

The conventional wisdom holds that our technological dependence began to form during a period of global climate change between three million and two million years ago, when Africa's woodlands transformed into savanna grasslands. Hominins, members of the human family, found themselves at a

IN BRIEF

A traditional view of human evolution holds that stone-tool technology originated with members of our genus, *Homo,* as an adaptation to shifting climate.

In this scenario, that adaptation quickly helped to establish a feedback loop that dramatically expanded brain size and technological prowess in our lineage.

Recently discovered stone tools from Kenya that date to 3.3 million years ago—long before the oldest known *Homo* fossils—have overturned this scenario.

© 2018 Scientific American

1

WORKERS DIG into the side of a hill at Lomekwi 3 in July 2016, looking for artifacts (*1*). They sift each bucket of sediment they remove, hoping to recover even the smallest fragments of interest (*2*). Every pebble is studied for signs of human modification.

crossroads. Their old food sources were vanishing. They had to adapt or face extinction. One lineage, that of the so-called robust australopithecines, coped by evolving huge molars and powerful jaws to process the tougher plant foods available in grassland environments. Another—the larger-brained *Homo*—invented stone tools that gave it access to a wide variety of food sources, including the animals that grazed on these new plants. With the rich stores of calories from meat, *Homo* could afford to fuel an even bigger brain, which could then invent new and better tools for getting still more calories. In short order, a feedback loop formed, one that propelled our brain size and powers of innovation to ever greater heights. By one million years ago the robust australopithecines disappeared, and *Homo* was well on its way to conquering the planet.

The Lomekwi tools have smashed that scenario to pieces. Not only are they too old to belong to *Homo,* but they also predate the climate shift that supposedly kindled our ancestors' drive to create. And without any cut-marked bones or other signs of butchery at the site, it is not at all certain that the tools were used to process animal foods. What is more, such a vast expanse of time separates the Lomekwi tools from the next oldest implements on record that it is impossible to connect them to the rest of humanity's technological endeavoring, suggesting the advent of stone tools was not necessarily the watershed moment experts have always envisioned it to be.

These new discoveries have scientists scrambling to figure out when and how our predecessors acquired the cognitive and physical traits needed to conceptualize and fashion stone tools and to pass their craft to the next generation. If multiple lineages made tools from rock, researchers will need to rethink much of what they thought they knew about the origins of technology and how it shaped our branch of the family tree.

Dawn breaks gently in the bush—a slow brightening of sky, a creeping swell of birdsong—and the team's campsite, on the bank of a dry riverbed about a mile from Lomekwi 3, comes to life. By 6:30 A.M. the workers emerge from their tents and head to the makeshift dining table for breakfast, walking along a gravel path lined with stones to deter the snakes and scorpions. Within the hour they pile into Land Cruisers and set off on a bone-rattling ride to the excavation.

The team is down one vehicle and short on seats in the remaining two, so archaeologist Hélène Roche has decided to stay at camp. Roche is an emeritus director of research at the French

Map by Amanda Montañez

© 2018 Scientific American

National Center for Scientific Research and an expert in early stone-tool technologies. She has short, sand-colored hair, and she dresses in desert hues. Her voice is low and crisp. Roche led the archaeological research in western Turkana for 17 years before handing the reins to Harmand and Lewis in 2011. She has returned for the second half of this expedition to see how they are faring. I remain at camp for the day to ask her about the history of work in this region.

"When I started in archaeology, we were just getting used to having stone tools at 1.8 [million years ago] at Olduvai," Roche recalls. In 1964 Kenyan paleoanthropologist Louis Leakey announced that he had found *Homo*-like fossils in association with what were then the oldest known artifacts in the world, stone tools from Tanzania's Olduvai Gorge (referred to as Oldowan tools). He assigned the fossils to a new species, *Homo habilis,* the "handy man," cementing the idea that stone toolmaking was linked to the emergence of *Homo*.

1

Hints that stone tools might have originated before *Homo* soon arrived, however. In the 1970s Roche, then a graduate student, discovered older Oldowan stone tools at a site in Ethiopia called Gona. When archaeologist Sileshi Semaw, now at the National Center for Research on Human Evolution in Burgos, Spain, and his colleagues eventually analyzed the tools, they reported them to be 2.6 million years old. Because no hominin remains turned up with the tools, researchers could not be sure which species made them. Semaw and his team proposed that a small-brained australopithecine species found at a different site nearby—*Australopithecus garhi*—was the toolmaker. Few were swayed by that argument, however. *Homo* was still the favorite candidate, even though, at the time, the oldest known *Homo* fossil was only 2.4 million years old. (A recent find has extended the fossil record of *Homo* back to 2.8 million years ago.)

Yet as old as they were, the Gona artifacts looked too skillfully wrought to represent humanity's first foray into stone-tool manufacturing. So did other ancient tools that began to emerge, including some from western Turkana. In the 1990s Roche found 2.3-million-year-old Oldowan stone tools at a site five miles from here known as Lokalalei 2c. She realized that in many instances, the site preserved entire knapping sequences that she could piece together like a 3-D puzzle. By refitting the Lokalalei flakes to the cores from which they were detached, Roche and her colleagues could show that toolmakers struck as many as 70 flakes from a single core. This impressive feat required an understanding of the rock shape best suited to flaking (flat on one side and convex on the other) and careful planning to maintain that shape while knapping. "You cannot imagine what it is like to hold the pieces together and reconstruct what [the toolmaker] has done and how he has done it, to go inside the prehistoric mind," she says.

It was becoming clear that the sophistication evident in the tools from Gona, Lokalalei and elsewhere could not have sprung fully formed from the minds of these knappers. Some kind of technological tradition must have preceded the Oldowan.

In 2010 far older signs of stone-tool technology came to light. Zeresenay Alemseged, now at the University of Chicago, and his colleagues reported that they had found two animal bones bearing what appeared to be cut marks from stone tools at the site of Dikika in Ethiopia. The bones dated to 3.4 million years ago, hundreds of thousands of years before the earliest known traces of *Homo*. The researchers credited the marks to *Australopithecus afarensis,* a species that was still apelike in many respects, with about as much gray matter as a chimpanzee has and a body that retained some adaptations to life in the trees—hardly the brainy, fully terrestrial hominin that researchers had traditionally expected the first butcher to be. The claims did not go unchallenged, however. Some experts countered that animals could have trampled the bones. Without the stone tools themselves, the critics argued, the Dikika scars could not qualify as tool-inflicted marks—and the question of just how far back in time technology originated remained unresolved.

Around the time the battle over the Dikika bones erupted, Harmand and Lewis began to hatch a plan to look for the older stone tools that the Dikika marks, along with the too-good-to-be-first tools from Gona and Lokalalei, implied should exist. In the summer of 2011 they set out in search of new archaeological sites on the western side of Lake Turkana.

The Turkana basin, as well as much of the Great Rift Valley in which it sits, is a paleoanthropologist's dream. Not only does it harbor an abundance of fossils and artifacts, but it preserves them in rocks that, with some sleuthing, can be dated with a relatively high degree of certainty. The region's history of volcanic eruptions and fluctuating water levels is recorded in the layers of sediment that have accumulated over eons to form a sort of layer cake. Water and wind erosion have exposed cross sections of the cake in locations throughout the basin. Tectonic activity has pushed some sections higher and other sections lower than they once were, but as long as any given exposure preserves at least a few layers of the cake, researchers can figure out where in the geologic sequence it comes from and thus how old it is.

© 2018 Scientific American

EXCAVATORS CHIPPED away at the sediments for weeks before finding any artifacts (*1*). The first finds were flakes produced incidentally during knapping (*2*). A volcanic ash layer called the Toroto Tuff helped to establish the age of the site (*3*).

To navigate the rough, roadless landscape, the team drives in the dry riverbeds, called *lagas,* that snake through the region, running from the lake to points west. On July 9, 2011, the researchers were headed to a site where, 12 years earlier, a different team had found a 3.5-million-year-old skull of another hominin species, *Kenyanthropus platyops,* when they took the wrong branch of the Lomekwi *laga* and got lost. Climbing a nearby hillside to get a better view of the terrain, they realized that they had ended up in just the kind of place that is promising for finding ancient remains. Outcrops of soft lake sediments, which tend to preserve fossils and artifacts well, surrounded them. And the researchers knew from previous geologic mapping of the region that all the sediments along this *laga* were more than 2.7 million years old. They decided to look around.

Within a couple of hours Sammy Lokorodi, one of the Turkana members of the team, found several rocks bearing hallmarks of knapping—adjacent, scoop-shaped scars where sharp flakes had been chipped off. Could these be the older, more primitive tools that the team was looking for? Maybe. But the tools were found on the surface. A modern-day human—perhaps a passing Turkana nomad—could have made them and left them there. The researchers knew that to make a convincing case that the tools were ancient, they would have to find more of them, sealed in sediments that had lain undisturbed since their deposition, and conduct detailed geologic analyses of the site to establish the age of the artifacts more precisely. Their work had just begun.

By the time the researchers went public with their discovery, describing it in 2015 in *Nature,* they had excavated 19 stone tools from a 140-square-foot area. And they had correlated the position of the sediment layer that held the tools to layers of rock with known ages, including a 3.31-million-year-old layer of compacted volcanic ash called the Toroto Tuff and a magnetically reversed layer from a time, 3.33 million years ago, when the earth's magnetic poles switched places. They had also located the source of the raw material for the tools—a 3.33-million-year-old layer of beach containing cobbles of volcanic basalt and phonolite, along with fish and crocodile fossils that show just how much higher lake levels were back then as compared with today. Together these clues indicated that the tools dated to a stunning 3.3 million years ago—700,000 years older than the Gona tools and half a million years older than the earliest fossil of *Homo.*

The artifacts have little in common with Oldowan tools. They are far larger, with some flakes the size of a human hand. And experiments indicate that they were knapped using different techniques. Oldowan knappers favored a freehand style, striking a hammerstone held in one hand against a core held in the other, Harmand explains. The Lomekwi knappers, in contrast, would either bang a core they held in both hands against an anvil lying on the ground or place a core on the anvil and hit it with a hammerstone. The methods and finished products demonstrate an understanding of the fracture mechanics of stone but show considerably less dexterity and planning than are evident in tools from Gona and Lokalalei. The researchers had found their pre-Oldowan stone-tool tradition. They call it the Lomekwian.

Not everyone is convinced that the Lomekwi tools are as old as the discovery team claims. Some skeptics contend that the team has not proved that the artifacts originated from the sediments dated to 3.3 million years ago. Discoveries made in the 2016 field season, including the knapping debris, as well as a handful of new tools recovered during excavation, may help allay those concerns. But even researchers who accept the age and the argument that the rocks were shaped by hominins are grappling with what the find means.

First, who made the tools? To date, the team has not recovered any hominin remains from the site, apart from a single, enigmatic tooth. The age and geographical location of the tools suggest three possibilities: *K. platyops,* the only hominin species known to have inhabited western Turkana at the time; *A. afarensis,* the species found in association with the cut-marked animal bones from Dikika; and *Australopithecus deyiremeda,* a species

© 2018 Scientific American

that was recently named, based on a partial jawbone found in Ethiopia. Either *K. platyops* or *A. afarensis* would be surprising because both those species had a brain about the size of a chimp's—not the enlarged brain researchers thought the first toolmaker would have. (*A. deyiremeda*'s brain size is unknown.)

Small brain size is not the only anatomical trait that experts did not expect to see in an ancient knapper. Paleoanthropologists thought that tool use arose after our ancestors had abandoned life in the trees to become committed terrestrial bipeds. In this scenario, only after their hands had been freed from the demands of climbing could hominins evolve the hand shape needed to make stone tools. Yet studies of *A. afarensis,* the only one of these three species for which bones below the head have been found, indicate that although it was a capable biped on the ground, it retained some traits that would have allowed it to climb trees for food or safety. Just how important was the shift away from life in the trees to life on the ground in the emergence of stone-tool technology?

The Lomekwi 3 tools are also forcing scientists to reconsider why hominins invented stone tools to begin with. Reconstruction of the paleoenvironment of the greater Lomekwi area 3.3 million years ago indicates that it was wooded, not the savanna experts thought had forged *Homo*'s stone-working skills.

THE LOMEKWI 3 TOOLS ARE ALSO FORCING SCIENTISTS TO RECONSIDER WHY HOMININS INVENTED STONE TOOLS TO BEGIN WITH.

Perhaps the biggest question: Why are the Lomekwi 3 tools so isolated in time? If stone-tool manufacture was the game-changing development that experts have always thought it to be, why did it not catch on as soon as it first appeared and initiate the feedback loop that expanded the brain?

Recent studies may help explain how a hominin more primitive than *Homo* could have come to make stone tools. It turns out that some of the differences in cognitive ability between hominins and other primates may not be as great as previously thought.

Observations of our closest primate cousins, for example, hint that even though they do not manufacture stone tools in the wild, they possess many of the cognitive abilities needed to do so. David Braun of George Washington University and Susana Carvalho of Oxford have found that in Bossou, Guinea, wild chimps that use stones to crack open nuts understand the physical properties of different rocks. The researchers shipped assorted stones from Kenya to Bossou and made them available to the chimps for their nut-cracking activities. Despite not having prior experience with these kinds of rock, the chimps consistently selected the ones with the best qualities for the job. And experiments with captive bonobos carried out by Nicholas Toth of the Stone Age Institute in Bloomington, Ind., and his colleagues have shown that they can be trained to make sharp flakes and use them to cut rope. "I have no doubt that our apes could replicate what [Harmand and her team] have at Lomekwi, given the right raw material," Toth asserts.

Even inventing stone tools in the first place may not have required special genius. In October 2016 Tomos Proffitt of Oxford and his colleagues reported that they had observed wild capuchin monkeys in Brazil's Serra da Capivara National Park unintentionally making sharp stone flakes that look for all the world like Oldowan tools. Quartzite cobbles abound in the monkeys' environment, and they will often pick up one cobble and bash it against another embedded in the ground that serves as an anvil. All the bashing dislodges sharp flakes that have the hallmarks of intentionally produced stone tools, including the scooplike shape that arises from what is known as conchoidal fracturing. The monkeys ignore the flakes, however. Instead they seem to be pulverizing the quartz to eat it—they pause between strikes to lick the resulting dust from the anvil. Perhaps early hominins invented their stone flakes by accident, too, or found naturally sharp stones in their environment, and only later, once they found a good use for them, began fashioning them on purpose.

The possibility that the Lomekwi toolmakers had hands that were at once capable of knapping and climbing in trees does not seem so improbable either, once one considers what our primate cousins can manage. The modern human hand, with its short, straight fingers and long, opposable thumb, is purpose-built for power, precision and dexterity—traits we exploit every time we swing a hammer, turn a key or send a text. Yet as the observations of chimps, bonobos and capuchins show, other primates with hands built for grasping tree branches can be surprisingly dexterous. The hands of partially arboreal hominins could have been similarly clever.

In fact, recent studies of the fossilized hand bones of three small-brained hominin species from South Africa—*Australopithecus africanus, Australopithecus sediba* and *Homo naledi*—show evidence for exactly this combination of activities. All three species have curved fingers—a trait associated with climbing. Yet in other respects, their hands look like those of toolmakers. Tracy Kivell and Matt Skinner, both at the University of Kent in England, examined the internal structure of the hand bones, which reflects the loading forces sustained in life, and found a pattern consistent with that seen in hominins known to have made and used stone tools and different from the internal structure of the hand bones of chimps. "Being a good climber and a dexterous toolmaker are not mutually exclusive," Kivell says. A variety of hand shapes can make and use stone tools, she explains. The changes the human hand eventually underwent just optimized it for the job.

Friday is *choma* night for the lomekwi team—roasted goat will be served for dinner. Nick Taylor of Stony Brook, a droll Brit, is taking advantage of the menu to try to figure out what purpose the Lomekwi stone tools served. This morning one of the local Turkana shepherds brought the purchased animal for slaughter. This afternoon, as the sun begins its descent and meal preparations begin, Taylor asks camp kitchen manager Alfred "Kole" Koki to try to process the carcass with replicas of the Lomekwi tools. Koki, an experienced butcher, doubts they will work. But he gamely takes a two-inch-long flake and starts slicing. He manages to skin most of the animal and carves some of the meat with the sharp-edged rocks, discarding them as they become dull, before reclaiming his steel knife to finish the job.

Taylor observes how Koki instinctively holds each flake and

© 2018 Scientific American

SONIA HARMAND and husband, Jason Lewis, co-direct the West Turkana Archaeological Project that discovered Lomekwi 3.

how long it retains its edge before Koki requests a new one. Taylor keeps the used replica flakes so that later he and his colleagues can compare their damaged edges with those of the real flakes. He will also collect some of the bones to study what kind of cut marks the carving might have left on them. And he will try using the tools to cut plant materials, including wood and tubers. In addition, Taylor is looking for any residues on the Lomekwi tools that might provide clues to what they were processing.

For whatever reason the Lomekwi hominins made stone tools, their tradition does not appear to have stuck. Nearly 700,000 years separates their implements from the next oldest tools at Gona. Perhaps hominins did indeed have a stone-tool culture spanning that time, and archaeologists have just not found it yet. But maybe the Lomekwi stone-working was just a flash in the pan, unrelated to the Oldowan technology that followed. Even the Oldowan record is patchy and variable, showing different tool styles at different times and places, without much continuity among them. As Roche puts it, "There is not one Oldowan but Oldowans."

This pattern suggests to many archaeologists that populations in multiple lineages of hominins and possibly other primates may have experimented with stone-tool production independently, only to have their inventions fizzle out, unbequeathed to the next generation. "We used to think that once you have toolmaking, we're off to the races," observes Dietrich Stout of Emory University. But maybe with these early populations, he says, technology was not important to their adaptation, so it simply faded away.

Around two million years ago, however, something changed. The tools from this period start to look as though they were manufactured according to the same rules. By around 1.7 million years ago a more sophisticated technology arises: the Acheulean. Known for its hand ax, the Swiss Army knife of the Paleolithic, the Acheulean tradition spread across Africa and into other parts of the Old World.

Braun thinks the shift has to do with improved information transmission. Chimps appear to have what he calls low-fidelity transmission of behavior based on observational learning. It works pretty well for simple tasks: by the end of his team's six-week-long experiment with the Bossou chimps, the entire community was using the rocks the same way. The activity seemed to spread by means of a recycling behavior in which one individual, typically a juvenile, would watch another, usually an adult, use a certain type of rock to crack nuts, after which the youngster would try to use the adult's tool set to achieve the same ends.

Modern humans, in contrast, actively teach others how to do complex things—from baking a cake to flying a plane—which is a high-fidelity form of transmission. Perhaps, Braun suggests, the variability seen in the Lomekwi tools and in those of the early Oldowan is the result of lower-fidelity transmission, and the standardization of the later Oldowan and the more sophisticated Acheulean signals the development of a more effective means of sharing knowledge, one that allowed humans to ratchet up their technological complexity.

As ancient as the tools from Lomekwi 3 are, the team suspects that even older ones are out there, awaiting discovery. One day, while the rest of the team is excavating, Lewis, Lokorodi and Xavier Boës, a geologist at the French National Institute for Preventive Archaeological Research, set out to look for them. They head for an area known to have sediments older than those at Lomekwi 3, speeding up the *laga* in a cloud of dust. They are taking the same branch they meant to take on that day in 2011 when they lost their way and discovered Lomekwi 3.

When they reach their destination, they fan out, eyes trained on the ground, scanning for signs of human handiwork in a sea of rocks baked red by the sun. Before long, Lokorodi spies cobbles bearing scoop-shaped scars. In theory, they could be more than 3.5 million years old. But the team will have to follow the same painstaking procedures it carried out at Lomekwi 3. The researchers will have to determine whether the rocks have been shaped by humans and, if so, figure out which stratigraphic level they eroded from, pinpoint the age of that level and then find more of them undisturbed in the ground. Lewis photographs the rocks and notes their location for possible survey in the future. The team will also explore a promising area about three miles from Lomekwi 3 that has sediments dating to more than four million years ago.

Figuring out what technology came before and after Lomekwi 3 and getting a clearer picture of how the environment was shifting will be critical to elucidating the correlations among dietary change, tools and the origins of *Homo*. "Maybe the links are all the same, but everything happened earlier," Lewis offers. "The pieces have exploded, but that doesn't mean they won't reassemble."

"We know quite a lot now but not enough," Roche says of the discoveries in western Turkana. "This is only the beginning." SA

Kate Wong is a senior editor at *Scientific American*.

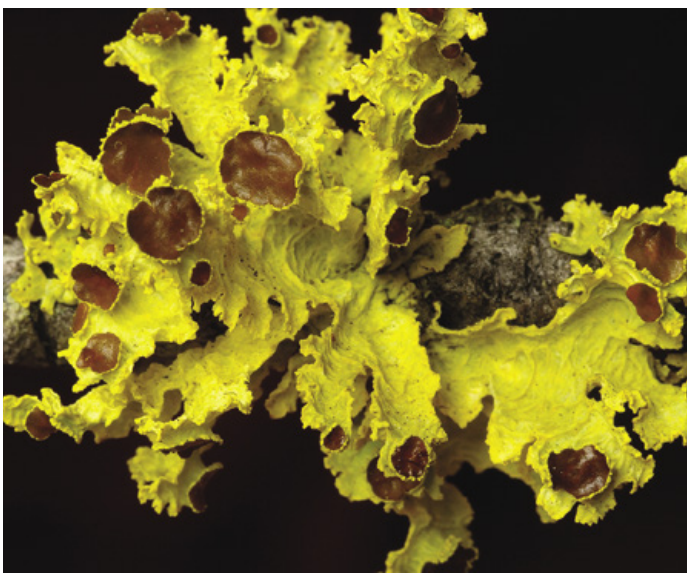
MORE TO EXPLORE

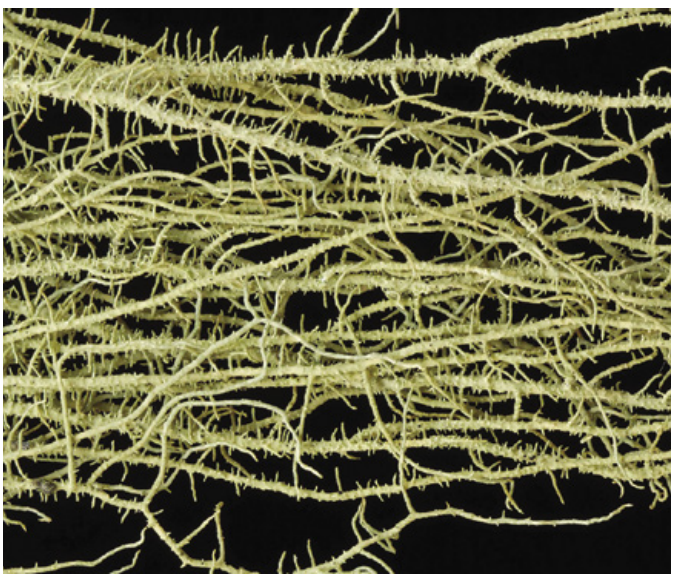
3.3-Million-Year-Old Stone Tools from Lomekwi 3, West Turkana, Kenya. Sonia Harmand et al. in *Nature*, Vol. 521, pages 310–315; May 21, 2015.

Wild Monkeys Flake Stone Tools. Tomos Proffitt et al. in *Nature*, Vol. 539, pages 85–88; November 3, 2016.

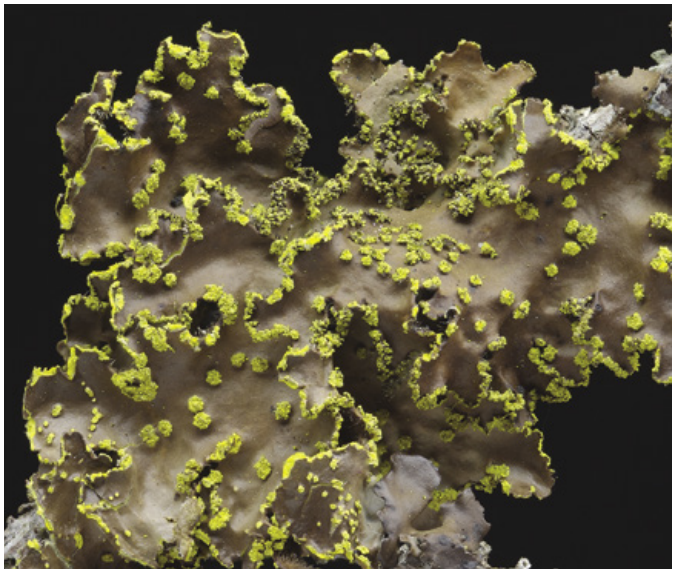
scientificamerican.com/magazine/sa

© 2018 Scientific American





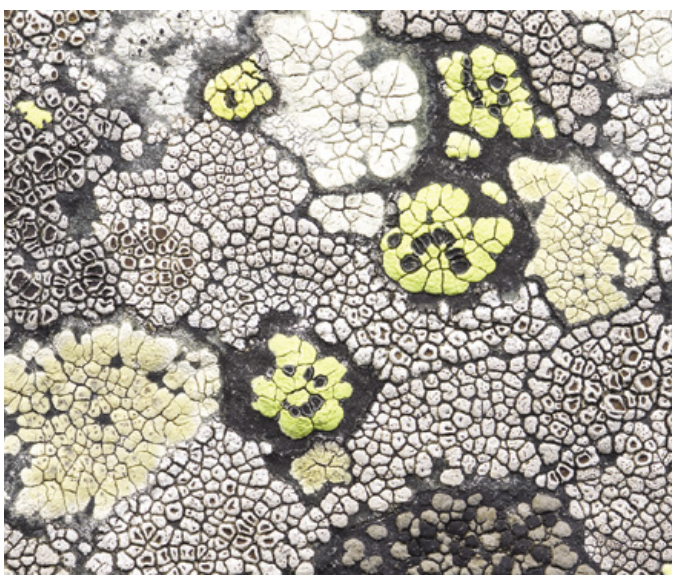
© 2018 Scientific American

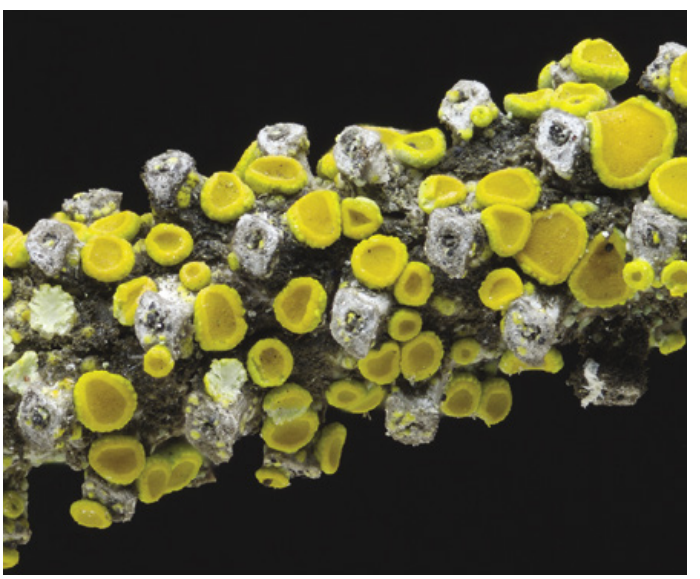


The Meaning of Lichen

How a naturalist's observations in the wilds of British Columbia inspired a scientist to discover hidden symbioses—overturning 150 years of accepted scientific wisdom

By Erica Gies





© 2018 Scientific American

TREVOR GOWARD LETS ME LEAD, SO WE TRAVEL through the mixed forest at my pace. This is a nod to his rangy 6'5" figure and the rapid strides he makes across barely discernible deer and bear paths on his land adjacent to Wells Gray Provincial Park in British Columbia. But mostly he is making space for my observations, my innate way of experiencing the landscape. What engages me? How do I see? I brake in front of a small, white-green growth on the trunk of an aspen. "Look there!" he says excitedly, inviting me to peer into the ragbag lichen through a magnifying lens. Suddenly I am in another world, looking down whorls studded with black dots into little caverns sprinkled with superfine dust.

Goward, white hair sticking up haphazardly, is wearing three flannel shirts on this crisp fall day. A hand lens hangs on a string around his neck, as an Australian shepherd named Purple trots along with him. He seems more mountain man than scientist, a naturalist in the tradition of Charles Darwin or Henry David Thoreau. Goward's scientific love is lichens—those growths that look like little mosses or colored crusts stuck to trees and rocks everywhere. He is inseparable from this place, where he has spent most of his adult life after growing up in a city south of the park. Now 64, he rarely leaves. "It has become my center of spiritual gravity," he tells me. It's not hard to see why. Most of the park has no road access and is rarely seen by humans. Wells Gray's 1.3 million acres were formed by volcanoes and glaciers; its river valleys, sheer rock mountains, alpine meadows and waterfall spray zones foster rich biodiversity. "I came to understand that the lichens here are pretty special," among the world's most diverse, Goward beams. There are hundreds of species and counting. His careful attention to this one place, like conservationist Aldo Leopold's beloved Sauk County, Wisconsin, allows him to see connections that others might miss.

Goward stumbled upon lichenology when he was educating himself about different branches of nature. "I made a point each year of learning as much as I could about a different taxonomic group. One year it was birds, then plants, then mushrooms, then insects." When he got to lichens, he was smitten. Since then, despite being self-taught, he has become the go-to expert in central British Columbia for everyone from atmospheric scientists to gold prospectors to caribou biologists. Several lichen species have been named after him. He has published three taxonomic guides to lichens and has earned a spot as an associate member of the University of British Columbia botany department.

And yet Goward is a maverick in the scientific world. His radical thought experiments about lichens, published in 12 provocative essays, available on his Web site, *Ways of Enlichenment,* have been both ridiculed and lauded—but largely ignored by most researchers because he holds no scientific degrees and

IN BRIEF

Trevor Goward, who has no scientific degree, has inspired scientists to upend the understanding of lichens by closely observing nature.

His insights, praised by some academics and discounted by others, are a strong reminder that biology, and science, may be getting too removed from the natural world, that mavericks can be brilliant, and that networks may be the most enduring lifeform, not individuals.

PRECEDING PAGES: TIM WHEELER

© 2018 Scientific American

NATURALIST Goward examines a ragbag lichen on an aspen. His deep observations have upset biology, like those of his predecessors Charles Darwin and Henry David Thoreau.

ERICA GIES

because many of his ideas are not supported by rigorous data. Still, Goward's astute observations and deep thinking follow in the footsteps of Darwin's and Thoreau's approaches—which, much more than laboratory science, formed the basis of the theories of evolution and ecology. People who are open to considering his ideas say they come away with mind-expanding food for thought about lichens, biology and all life. Goward's longtime friend and sometimes co-author Toby Spribille, a lichenologist at the University of Alberta, says Goward's essays contain many gold nuggets: "Frankly, I think they are brilliant."

In the forest, Goward exudes a quiet, ebullient joy, the yang to his yin: a dark, realistic assessment of humanity's folly. When we stop, he leans on a sturdy walking stick and delivers extended soliloquies about how elements of the ecosystem interact. Reading the lichens informs him about soil chemistry, rain patterns and plant nutrients. He shows me a species growing on a hemlock, unusual because conifer bark is usually too acidic to support these kinds of lichens. So why are they there? In a 2000 paper Goward and his co-author André Arsenault found that the answer lies in a mature trembling aspen nearby. Water dripping from its branches becomes a leachate, which, when it falls onto the conifer's bark, lessens the acidity, allowing the lichen to thrive. They dubbed this interaction the drip-zone effect.

Goward learns from every life-form, including Purple, who waits on us patiently when she is not conducting her own observations: Scat from a pine marten. Red squirrel chatter! Although Goward knows French, Latin, and some conversational German and Swedish, he remarks that "mostly these days I speak lichen and maybe a bit of dog." He says he can learn from Purple's way of seeing. That may seem eccentric, but Goward respects First Nations peoples' ways of knowing, and learning from animals is a storied human tradition.

Modern science tends to ignore outsiders. But reductionist science is not the only way of knowing things. Naturalists were the forebears of science. Humans once lived much closer to the land and were keen observers who had a deep understanding of nature's interactions. Today biology tends to be focused on molecules, and failure to look up from instruments in the lab and actually observe how pieces interact in the natural world sometimes undermines discovery. A clinical focus can lead scientists to miss big-picture connections, such as an emerging understanding that networks may be a more enduring life-form than individuals. Indeed, it was an idea from Goward that inspired Spribille's lab work while he was a postdoc at the University of Montana. That work paid off with a major advance: a July 2016 cover story in *Science* that rocked the ossified world of lichenology. The discovery called into question the very nature of the lichen symbiosis, shedding fresh light on how symbioses across biology work, how natural selection proceeds and even how to define life-forms.

Lichens are both ubiquitous and fascinating. Perhaps more than 500 million years old, they occur on every continent and

© 2018 Scientific American

can thrive in some of the most inhospitable places on earth. They even survived for a year and a half in space, fully exposed to cosmic radiation, ultraviolet irradiation and vacuum conditions. The approximately 14,000 species of lichen come in a variety of forms: flat rounds on stones, scalloped leaves nestled among mosses, crusts clinging to tree bark, flowing strands hanging from branches, tiny trumpets tipped in red.

For centuries people thought they were plants (and then fungi). Then, in the 1860s, Swiss botanist Simon Schwendener discovered that they were a partnership between a fungus (an organism classified in its own kingdom because, unlike plants, it cannot make its own food) and an alga, an organism that feeds itself with photosynthesis but lacks the roots and stems of plants. The fungus apparently provided the structure of the lichen, and the alga provided food for the fungus via photosynthesis. (Later it was discovered that in some lichens, a cyanobacterium provided the food—and a handful of species contained both an alga and a cyanobacterium, along with the fungus.) Schwendener's discovery, at first resisted by the scientific community, ultimately made lichens the poster children for symbiosis, a mutually beneficial interaction among organisms. Since then, science has found symbioses across nature, including among the trillions of nonhuman microbes that cling to the scaffold of our bodies.

CLOSE EXAMINATION has revealed that *Bryoria fremontii* is not simply a partnership between a fungus and an alga, as long thought; a yeast is also involved.

Science over the past two centuries has largely viewed molecules, cells and species as individuals. Symbiosis challenges that notion. "Within a lichen," Spribille says, "algal cells and fungal cells may experience each other as individuals, but together they form a lichen that the feeding caribou sees as an individual: tasty." Natural selection happens on both scales simultaneously. Just as light is both a wave and a particle, the fungus and alga are both individuals and parts of a whole. Science's reductionist focus has made it nearly impossible to fully understand symbiosis, Spribille says. "Ecology was supposed to be the science of natural process and synthesis, but its backbone is severely strained under the mathematics of individuality."

In July 2016 Spribille and his co-authors took a major step forward in that understanding. Their big reveal in *Science:* many lichens have a second fungus in the house.

At the heart of their study is a pair of lichens to which Goward had drawn Spribille's attention: *Bryoria fremontii,* which is hairlike and often brown and eaten by northwestern indigenous peoples, and a similar lichen, *Bryoria tortuosa,* which is often a yellowish green and is toxic, with high levels of vulpinic acid. The two posed a fascinating conundrum. Despite their differences, a genetic analysis published in 2009 by Saara Velmala of the University of Helsinki and her colleagues, on which Goward was a co-author, showed that the two species consisted of the same fungus and same alga. Spribille recalled how this perplexing finding infected them both. "[Goward] took the question of how could these two different lichens—one of which is toxic, for God's sake—be identical." The question would not let go of Goward. And when Goward wrote about it, "by extension, it wouldn't let go of me."

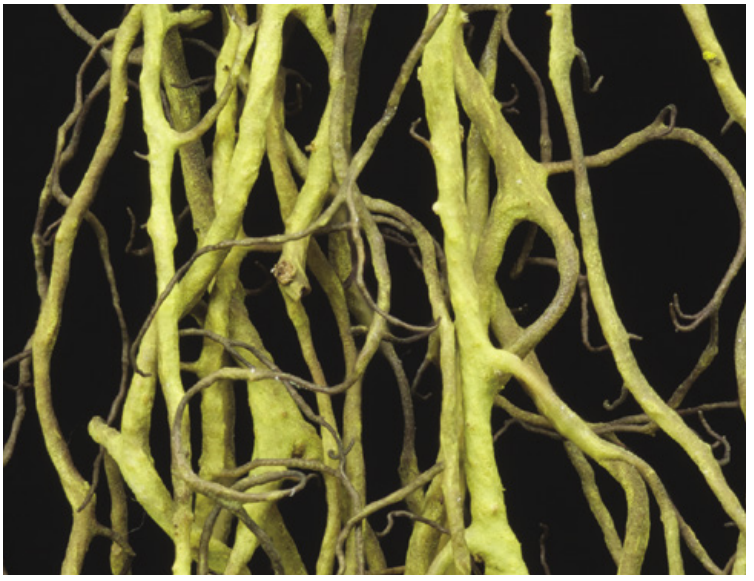
Aside from their different appearances and levels of vulpinic acid, Goward observed that the two lichens also had slightly different ecologies. Although they grew in some of the same places, *B. tortuosa* was found only on the summer-dry fringes of *B. fremontii*'s larger territory. In 2009 he proposed that lichens are formed not by the shape of their fungal partner but by a series of decisions made during the developmental dance between fungus and alga. One lichen can look different from another that is composed of the same partners because it took different turns during development. Goward suggested that the difference between the two species of *Bryoria* might stem from each of them having a different relationship with a third lifeform, a bacterium.

After five years of work in the lab, Spribille and his colleagues discovered that both *Bryoria* species did include a third partner. But it was not a bacterium; it was another fungus, known as a basidiomycete yeast. The toxic *Bryoria* contained a lot more of the yeast than the edible one. The team also demonstrated that the yeast was not a contaminant but had evolved with the other partners for more than 200 million years. Expanding their search to lichens across the globe, they found the yeast in 52 other sets (genera) of lichen. The finding dramatically expanded the world's understanding of lichens, opening the door to other insights. "Only now are we beginning to see that lichens really have pulled off a rare feat in evolution: a large multicellular organism but built entirely of microbes—and here's the amazing thing—without a scaffold," Spribille says. "Self-assembling, self-replicating, generation after symbiotic generation."

Goward first became interested in *B. fremontii* and *B. tortuosa* when he read ethnobotanist Nancy Turner's 1977 paper about *B. fremontii*'s importance to First Nations peoples. She said that women elders could easily distinguish the edible from the nonedible lichens. Although the two can have different col-

THIS PAGE AND OPPOSITE PAGE: TIM WHEELER

© 2018 Scientific American



BRYORIA TORTUOSA **hosts the same fungus and alga as the edible *B. fremontii*, but it has a much higher concentration of yeast, and it is poisonous.**

oring and a slightly different shape, they can also look quite alike. Elders use clues such as location, color and the types of neighboring lichens to tell them apart. When Stuart Crawford, a friend of Goward's with a degree in ethnobotany, showed bundles of the two lichens to an elder and conservationist from the Neskonlith band, the late Mary Thomas, she correctly identified the edible one every time.

Local people's wisdom does not always jibe with scientific explanations, Crawford says, but the result, based on observation, is correct. The locals told Crawford that they wait for *B. fremontii* to "ripen" on the tree. In fact, lichens do not ripen as do fruits and vegetables, but the darker color and its growth pattern on trees help the people distinguish it from its poisonous twin. These other realms of knowledge about *Bryoria* would have added interesting context to the *Science* paper, Spribille says, but "it didn't fit the word limit."

Three months after the paper was published, Crawford, who knows Spribille through Goward, got around to telling him something amazing. For years Crawford had been collecting writings from around the world—ancient Egypt, modern Mexico, medieval Russia, the Biblical Middle East, a European cookbook from the 1950s—of people using lichens to make bread and alcoholic beverages. In some cases, they were using them explicitly for leavening and fermentation. On some level, Crawford realized, people knew that lichens contained yeast or functioned like yeast. When he was working on his master's degree in Victoria, B.C., Crawford discussed the notion with a local microbrewer, who told him, "If you can figure out the recipe, I'll do a batch of beer" with it.

Spribille's openness to Goward's unconventional ways of thinking is perhaps a reflection of his hard-won path to science. He grew up in a fundamentalist Christian family in northwestern Montana, where his parents pulled him out of school after fourth grade to protect him from "the influences of the world." Spribille is telling me this via Skype from Austria, where it is late at night and his wife and young daughter are sleeping. His rectangular glasses frame blue eyes that frequently squeeze shut while he is talking, as if communicating with me is a little painful.

Circumstance could not restrain Spribille's intellectual curiosity. Intrigued about organisms he saw in the wild, he sought answers from biologists at a local U.S. Forest Service outpost. Eventually they recommended him for a job surveying vascular plants, and he could call up professors and authors with his burning questions. Goward was on Spribille's call list. "Trevor kept me on for two and a half hours," Spribille says fondly. That was more than 20 years ago. They have co-written several papers, and "we still haven't run out of things to talk about." Early on, Goward told Spribille that he had ideas that would turn lichenology upside down. "He said that I was delusional," Goward recalls. "But he wanted to hear the ideas."

Ultimately Spribille felt a keen desire for a formal education. He took the high school equivalency examination and found an opportunity to go to college in Germany. He later earned a Ph.D. in lichenology at the University of Graz in Austria and in 2017 began his new appointment as assistant professor of the ecology and evolution of symbiosis at the University of Alberta. During his postdoc at the University of Montana, he met John McCutcheon, one of the co-authors on the *Science* paper and head of the lab in which the work was done. McCutcheon credits the breakthrough to technological advances that allowed the researchers to find the tiny yeast and to cooperation among diverse co-workers. But also critical, he says, was Spribille's ability to look beyond what was assumed to be true. The human mind's tendency to restrict itself is part of what kept this yeast hidden for so long, he says: "When you're used to thinking there's just one fungus there, that's what you see."

THE UNIT OF LIFE MAY NOT BE AN INDIVIDUAL BUT A NETWORK, WHETHER AMONG THE ORGANISMS MAKING UP A LICHEN OR THE MICROBES OF THE HUMAN MICROBIOME.

Spribille, in turn, credits Goward with having "a huge influence on my thinking. [His essays] gave me license to think about lichens in a way that was not orthodox and freed me to see the patterns I worked out in *Bryoria* with my co-authors." Yet even with that, Spribille says, "one of the most difficult things was allowing myself to have an open mind to the idea that 150 years of literature may have entirely missed the theo-

© 2018 Scientific American

retical possibility that there would be more than one fungal partner in the lichen symbiosis."

While he appreciates his education, Spribille maintains that academia's emphasis on the canon of what others have established as important is inherently limiting. "You have this culture of prepared minds that makes it extremely difficult to think outside the box," Spribille says. "It creates the box."

That sounds plausible to Jonathan Foley, executive director of the California Academy of Sciences in San Francisco, who has a Ph.D. and had an acclaimed academic career. (Foley serves on *Scientific American*'s board of advisers.) When it comes to ideas, "the ivory tower is now an ivory fortress," he says. Academic culture's incentives to publish in accepted journals, get funding and obtain tenure are "not aligned with being wildly creative." After Sputnik, science became hyperprofessionalized, Foley says—"kind of Science, Inc. I think we lost part of our souls." The extreme specialization required for the biotech-heavy, molecule-focused world of biology today eliminates time to study taxonomy or epistemology. "There are people getting degrees in biological sciences at the best universities in America today who don't know the names of anything outdoors, who have never studied anything larger than a cell," Foley notes. That means a lot of biologists are lab-bound and rely on people like Goward to find the species they would like to study or even to suggest ideas for studies.

Also worrisome to Spribille is that his own students are petrified of being wrong, a psychological state incompatible with breakthroughs. For a counterexample, he points to Goward. In the case of *Bryoria,* Goward surmised that a third partner was present, although he incorrectly thought it was a bacterium. But being correct "is not the criterion of a brilliant mind," Spribille says. On the contrary, brilliant minds are characterized by indefatigable curiosity and questioning, traits that Spribille tries to encourage in his students. "I tell them, 'Just put all the ideas out there. Nobody here is going to make you feel bad about throwing out an idea that we may then not use.' I live by that."

Some of the most serious problems science is trying to solve today—climate change, loss of biodiversity, food and water security—require big, integrated views from multiple perspectives. Stepping out of the lab and back into nature to observe how natural systems actually work is a critical first move. One biologist at the University of the South challenged himself to try it. David George Haskell spent a year sitting in a square yard of old-growth forest in Tennessee, just observing, and wrote a Pulitzer-nominated book about it, *The Forest Unseen*. The experience was profoundly humbling, he says. "You wake up to the extent of your own ignorance. I'd been through decades of training and teaching as a biologist and had published scientific papers and so on, and sitting down in the woods, I realized I know so, so little about this place." From that humility sprouted seeds of curiosity and dozens of questions about relationships among plants and animals, their ecological history, and how that related to climate and geology. Haskell is now an adviser to the New York City–based Open Space Institute, helping it to identify lands for conservation that are most likely to be climate-resilient.

If knowledge comes mostly through reading scientific literature, "we're several steps removed from the actual phenomena we're discovering," Haskell explains. And while instruments are important to help scientists understand the world, "our bodies come preinstalled with all these amazing apps, and they connect directly into our consciousness," he says. "Through literally coming back to our senses, we can learn so much about the world."

LICHENS GROW on every continent, and they survived a year and a half on the outside of the International Space Station, fully exposed to cosmic radiation.

Goward has turned this ethic into a way of life. His house, named Edgewood Blue, on 10 acres abutting Wells Gray, has running water for a shower and sinks but no toilet. As I put on my coat and shoes one evening to head to the outhouse, Goward's partner, Curtis Björk, a botanist, encourages me to look up and admire the Milky Way, vivid in the lack of light pollution. When I ask why he and Björk have no toilet, Goward says they appreciate being forced to go outside every day, even in the depths of winter. On trips to the loo he has seen the Northern Lights and passing moose. When I jokingly whimper about getting wet or cold or chomped by summer mosquitoes, or even stalked by the cougar that recently swiped a neighbor's pigs, Goward is unapologetic: "That's real. Life isn't always comfortable."

To Goward, the real danger lies in separating ourselves from the natural world, living ensconced in cities, ignorant of how badly we are degrading nature. Haskell agrees and points out that this separation has ethical implications. "Trees, fungi, salamanders … these are our blood kin, if you believe Darwin." When we do not know the world, we have an imperfect sense of right and wrong in how our own behavior impacts ecosystems, he says.

But for scientists who may not have a year, or 30, to spend

THIS PAGE AND OPPOSITE PAGE: ERICA GIES

© 2018 Scientific American

AT HIS WOODLAND HOME, Goward hosts an ongoing parade of biologists, poets and astrophysicists who explore ideas about exotic life-forms and the pitfalls of humans separating from the natural world.

contemplating the wild, collaborations or friendships with people outside the academy or from different disciplines can open space for new discoveries, as they did for Spribille.

The decor at Edgewood is dominated by books, which serve as de facto wallpaper, lining homemade shelves in most rooms. The kitchen is Björk's fiefdom, and when dinner is ready, he sends Purple to fetch the humans. Purple eats at the table with us, displaying excellent manners.

Also served at the table are wide-ranging discussions. "We try to make this a place where anyone can express their ideas," Goward pronounces. The strengths and failings of modern science are a frequent theme, driven primarily by deep affection for it. Life and human relationships also take the spotlight, sometimes with quirky analogies to lichens. Although Goward is confident—sometimes bordering on arrogant—about his own ideas, he is eager to consider new information. His dialogue is peppered with references to authors. When I mentioned Haskell's book and an essay by the late Ursula K. Le Guin, he had read them by the next time we talked.

With the scientific world often reluctant to publish him, Goward spreads his ideas one person at a time. He and Björk host an ongoing parade of biologists, aspiring naturalists, poets, geographers, ecologists, astrophysicists and journalists who stay for a day or a week or longer in return for doing a bit of work at Edgewood. Both Spribille and Crawford have become regular guests. "I've done a lot of landscaping on his place," Crawford says proudly. "We have great and intellectually stimulating conversations."

Goward would like to create a more formal venue for learning, to increase "biological literacy" in the next generation, and has offered half of his land as a research center to Thompson Rivers University in nearby Kamloops, where he grew up. He also periodically invites people from various disciplines to meet for a few days of discourse.

On my visit Goward delves into one of his pet lines of inquiry: What are lichens, really? Are they organisms? Fungal greenhouses? Algal farmsteads? Ecosystems? Networks?

What you think lichens are might depend on your perspective. Because lichens have the scientific names of their fungi, that can create an implicit bias that the fungus is in charge, a limited perspective that Goward admits to having once upon a time. Today he sees lichens as a kind of koan. "The lichen by its very nature exists at a portal, a doorway," he says. "If you look in one direction, it's an organism. If you look in the other direction, it's an ecosystem." Goward's essays argue for seeing lichens not as their fungal or algal parts or even as ecosystems or organisms. Rather they are all these things, biological systems encapsulated in a membrane: lichens as emergent property. After all, the lichens that were sent into space survived when their algae alone did not.

Thinking of lichens as systems fits with a larger shift in biology from viewing the fundamental unit of life as the individual to that of community or partnerships. "Whether it is the microbiome within the human body or trees interacting with fungal partners belowground or lichens … we're seeing that networked relationships are more fundamental and persist longer within biological systems than individuals do," Haskell says.

To Goward, lichens are the organisms that are most obviously about relationships. As such, they provide insights into all of life. "Lichens are my window," he says, "but it's the act of looking at the world that's the interesting thing." Systems only hold together in the long term if the parts consider themselves integral to the whole and if the whole protects the parts, as lichens do. "That's what's going wrong with us," he says. "As individuals, we're not concerned with the whole." SA

Erica Gies writes about science and the environment from Victoria, B.C., and San Francisco. Her work appears in the *New York Times*, the *Guardian*, the *Economist*, *Ensia*, *National Geographic* and elsewhere.

MORE TO EXPLORE

Cyanolichen Distribution in Young Unmanaged Forests: A Dripzone Effect? Trevor Goward and André Arsenault in *Bryologist*, Vol. 103, No. 1, pages 28–37; Spring 2000.
The Forest Unseen: A Year's Watch in Nature. David George Haskell. Viking, 2012.
Basidiomycete Yeasts in the Cortex of Ascomycete Macrolichens. Toby Spribille et al. in *Science*, Vol. 353, pages 488–492; July 29, 2016.
The Songs of Trees: Stories from Nature's Great Connectors. David George Haskell. Viking, 2017.
Trevor Goward's essay series at his Ways of Enlichenment site: www.waysofenlichenment.net/ways/readings/index

scientificamerican.com/magazine/sa

© 2018 Scientific American

THE PERFECT BEAST

Mixotrophs, tiny sea creatures that hunt like animals but grow like plants, can change everything from fish populations to rates of global warming

By Aditee Mitra

Illustration by Mark Ross Studios

© 2018 Scientific American

SUCKER PUNCH: One mixotroph, *Dinophysis* (*right*), sucks photosynthesizing organs from another, *Mesodinium*.

© 2018 Scientific American

SUMMER SUNLIGHT FLICKERS THROUGH WARM WATERS OFF THE COAST OF Spain. The sea looks calm and peaceful. Near the surface, invisible to the naked eye, a swarm of microscopic plankton, some orange-pink and others dark green, swims in lazy circles, capturing the sun's rays and using the solar energy to make nutrients through photosynthesis. Suddenly, a tentacled creature called *Mesodinium*—at 22 microns, a giant next to some of the three-micron sun-gathering plankton—comes zigzagging through the waters, drawn by sugars and amino acids leaking from the smaller organisms. Its tentacles shoot out and engulf the hapless green prey, or nanoflagellates, which are completely consumed and digested.

The predator is pickier, though equally brutal, with its pink-colored prey, called cryptophytes. While it digests and destroys most pieces, the attacker takes in the organelles responsible for photosynthesis whole. Within minutes the pale *Mesodinium* starts to turn darker red as it fills its body with the stolen parts—chloroplasts and nucleosomes—which remain intact and functional. *Mesodinium* is not able to take in and use carbon dioxide as true photosynthesizers are, so it relies on its victims' chloroplasts to accomplish that task. The creature's dual strategy of hunting for food like an animal and photosynthesizing like a plant is known as mixotrophy.

Mesodinium does not, however, get to hold onto its loot for long. Nearby lurks yet another mixotroph, slightly larger and with different hunting skills: the dinoflagellate *Dinophysis*. It starts circling the other creature and then shoots out harpoon-like threads that immobilize the *Mesodinium*. The captor then administers the coup de grâce. It lances its target with an appendage called a peduncle, something that looks and acts like a straw, and sucks out the innards, including the stolen chloroplasts. These now third-hand photosynthetic factories are assimilated into the new host and begin churning away within the *Dinophysis,* giving it energy for life. Remnants of the original thief, the butchered *Mesodinium,* drift away.

These single-celled killers are only two examples of countless billions of mixotrophic plankton that swim in our seas. For a long time, most marine scientists dismissed mixotrophs as minor curiosities, compared with the two main groups of single-celled plankton that are supposed to dominate the base of the ocean's food web. One of these groups, plantlike phytoplankton, uses light energy and inorganic nutrients such as nitrates to proliferate. The other group, animallike zooplankton, eats these phytoplankton. In this way, nutrients are passed up the food

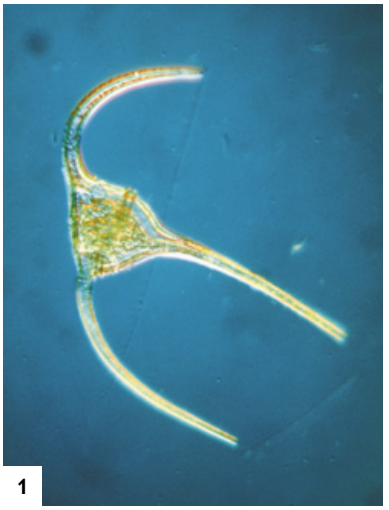
IN BRIEF

The ocean food web—key to the earth's ecosystem—has been thought to depend on two groups: plantlike and animallike plankton.

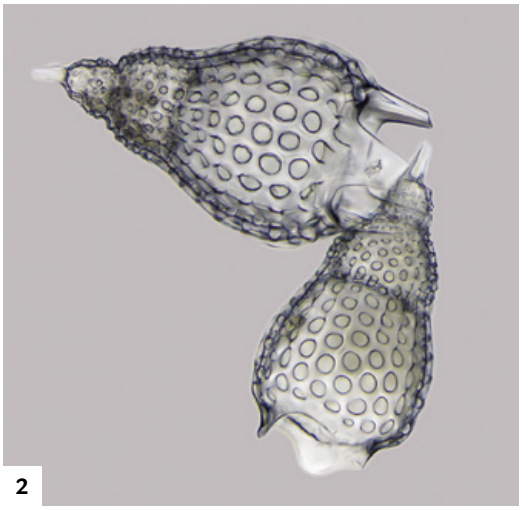
But new evidence shows many plankton are "mixotrophs" that use solar energy like plants yet hunt and kill prey to live.

These hybrid creatures have giant effects on global carbon levels, fish populations and harmful algal blooms.

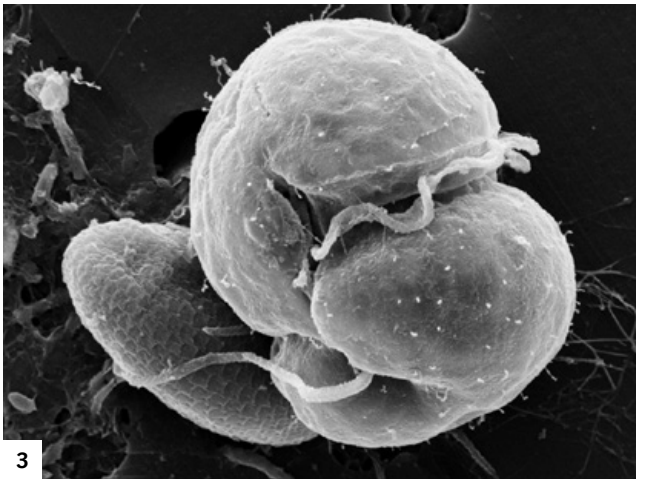
© 2018 Scientific American



1



2



3

TINY MONSTERS: **Mixotrophs come in several varieties. Some, such as *Tripos longipes* (*1*), can photosynthesize on their own and eat prey. But members of the order Nassellaria (*2*) steal photosynthetic organs from plankton victims. Species of *Karlodinium* (*3*) act as *T. longipes* does.**

web to bigger animals. Next to these two purists, mixotrophs were deemed to be inefficient oddballs, jacks-of-all-trades but masters of none. (There are some rare examples of mixotrophs on land, such as the carnivorous Venus flytrap plants.)

This traditional view of the ocean food web is wrong. Through experiments, observations and models of plankton populations, my colleagues and I have recently uncovered evidence that most single-celled plankton are neither purely plantlike nor pure plant eaters. The great majority are, in fact, mixotrophs. This means the bottom of the food web—and thus everything above it—is not controlled the way we thought it was. If most plankton are really mixotrophs, their numbers are not sharply limited by photosynthesis but can increase through eating. And solar energy, when available, can give an extra boost to eating-driven growth. These abilities have a ripple effect on many things, from the atmosphere to fish populations. Greater mixotroph activity, for instance, affects how quickly the oceans remove climate-warming carbon dioxide from seawater and air and lock it away in sea-bottom sediments. Mixotroph groups may be less susceptible to the waxing and waning of sunlight as seasons change. Such versatility and resilience have advantages. Populations of beneficial mixotrophs can provide sustenance for more fish larvae and increase the human food supply. But there can be negative consequences, too. Some mixotroph species cause harmful algal blooms that close shellfish hatcheries and lead to widespread fish kills.

If this new view of marine biology holds water—as recent findings by ourselves and other scientists have shown—it means the idea of an ocean ecology based on "plantlike" and "animallike" plankton is no longer tenable. There is a different, strange and powerful player amid the waves.

ATTACK OF THE TRIFFIDS

MIXOTROPHS are like something out of science fiction. The harpoons and peduncles of *Dinophysis* seem, on a smaller scale, like features of the planet-conquering plants in the famous 1951 sci-fi novel *The Day of the Triffids,* by John Wyndham. Triffids could use their roots to extract nutrients from the soil and to walk around. But they also had venomous stingers that they used like whips, blinding or killing humans and then feasting on the decomposing bodies.

The first time I heard of plankton mixotrophs was a little more than a decade ago, during my Ph.D. research on microzooplankton, which are smaller forms of animallike plankton. (The smaller plantlike organisms, likewise, are called microphytoplankton.) Textbooks described mixotrophs as a weird oceanic rarity. Yet with their dual ability to photosynthesize and hunt, combined in a single cell, mixotrophs appeared to be nature's perfect beasts. Because evolution tends to favor efficiency, it surprised me that mixotrophy was not more common. Searching for more information about it, I found a whole suite of exciting publications by Diane Stoecker, a plankton ecologist at the University of Maryland's Horn Point Laboratory, whose field and laboratory work suggested that mixotrophy existed among ocean plankton. I got in touch with Stoecker, and our discussions convinced me that more mixotrophs were out there. But how many, and what were they doing?

My own expertise is in building mathematical models of food webs to understand the behavior of different organisms in them. In essence, these are simulations that run on a computer. In searching through ocean ecosystem models, I could not find one that simulated the details of mixotrophs' double lives. Nor could I find funding for a project to build one; scientific review committees did not think mixotrophs were very important. So I spent my days working as a biodiversity officer for the local government in Bridgend, Wales, and in the evenings I worked on

ERIC GRAVE *Science Source (1)*; FRANK FOX *mikro-foto.de (2)*; VINCENT LOVKO *Mote Marine Laboratory (3)*

© 2018 Scientific American

the model with my marine biologist husband, Kevin Flynn. By the spring of 2009 we had a working simulation, one that could represent different populations of mixotrophs, some hunting more and some doing more photosynthesis. It was published that year in the *Journal of Plankton Research*.

Our goal was to show that a model including dual-natured plankton could simulate marine ecology more realistically than other models that segregated ocean populations into predators and plants. We changed mixotroph characteristics in the model until we had simulations that captured real-life observations of nutrient flow within food webs, as well as interactions among other plankton types such as bacteria and tiny crustaceans called copepods. These food web dynamics, which we published in 2010 in the *Journal of Marine Systems,* were very different from the segregated plankton models.

We needed to go beyond computer simulations, however. We had to gather evidence supporting our hypothesis that mixotrophs were critical drivers of the ebb and flow of nutrients through all parts of the ocean and all creatures in it. This time there was money. Impressed by the success of our models, a foundation called the Leverhulme Trust supported a series of meetings in Europe and the U.S., where for the first time, researchers who worked with mixotrophs in the field and the lab could share what they knew.

THESE SCUMLIKE BLOOMS ARE NOT TOXIC, BUT THEY DO BLOCK SUNLIGHT, PUTTING A CRIMP IN A NUTRIENT CYCLE THAT FEEDS TINY FISH LARVAE. FEWER BLOOMS MEAN MORE FISH.

A WORLD OF HYBRIDS

AT THE FIRST MEETING in 2011, our group (we dubbed ourselves "Team Mixotroph") made a list of all the plankton species known to both hunt and photosynthesize. Scientists had, over the course of decades, identified mixotroph species in water samples across all marine systems, from coasts to midocean areas and from the poles to the equator. They carried out experiments in labs onboard the research ships. And in some instances, they brought back the plankton to their institutes and carried out additional experiments using different nutrients, prey or light intensities to find out how the mixotrophs behaved under different environmental conditions. Until our group started combining these observations, most researchers thought they were looking at small and unusual occurrences rather than a common life-form throughout the world's oceans.

Bringing these data together drove us to the conclusion that there was a lot of mixotrophy happening in the sea and that it was ecologically essential. For example, Per Juel Hansen, a plankton ecophysiologist at the University of Copenhagen, and his colleagues demonstrated that without sufficient cryptophyte prey (those pink plankton mentioned earlier), the *Mesodinium* population would not be able to acquire stolen chloroplasts and would die out. Stoecker and her team, as well as Hae Jin Jeong and his colleagues at Seoul National University, showed that when mixotrophs were actively photosynthesizing, they ate other plankton at higher rates than those that were not performing photosynthesis: one mode enhanced the other. And when light and nutrients were abundant, these particular mixotrophs had a much higher growth rate than did plankton stuck with only a single process.

In 2012 we started looking beyond the mere presence of mixotrophs in marine ecosystems and began to identify different groups based on what they eat, how they eat and how they photosynthesize. It turns out there are four distinct types, and each occupies a different place along a spectrum of mixed behavior.

The first criterion for discriminating among different types of mixotrophs was to identify the source of their photosynthetic abilities. Did they have an inherent capacity for using light to make food, or did they have to attack and steal photosystems from their prey? We called the group with their own genetic capabilities to produce and maintain the body parts for photosynthesis constitutive mixotrophs. This group includes a lot of benign and ecologically important creatures that are critical links in ocean food chains. But it also contains plenty of troublemakers. We see their effects in destructive algal blooms when their populations get out of hand. For example, the mixotroph *Karlodinium* is renowned for causing massive fish kills across the world, from the Chesapeake Bay to Malaysian coastal waters. *Prymnesium,* another constitutive mixotroph that has caused similar fish kills off the coast of Texas and in the backwaters of the Norfolk Broads in England, releases a chemical that destroys the integrity of cell membranes belonging to competing plankton. As a result, these plankton swell up and explode. Then *Prymnesium* consumes the debris. Toxins produced by another species, *Alexandrium,* make their way into mollusks as those animals ingest plankton-laced water. Entire oyster, mussel and clam fisheries are then closed because humans who consume seafood contaminated with *Alexandrium* can get paralytic shellfish poisoning.

The second group does not have inherent photosynthetic capabilities and has to hijack them. These are called nonconstitutive mixotrophs and include *Mesodinium* and *Dinophysis.* This group is a big collective. Its members' use of photosynthesis had been considered a supplemental survival mechanism employed only when there was not a lot of prey around. Now we realize that they harness solar energy much more often, and it is usually a crucial part of their way of life.

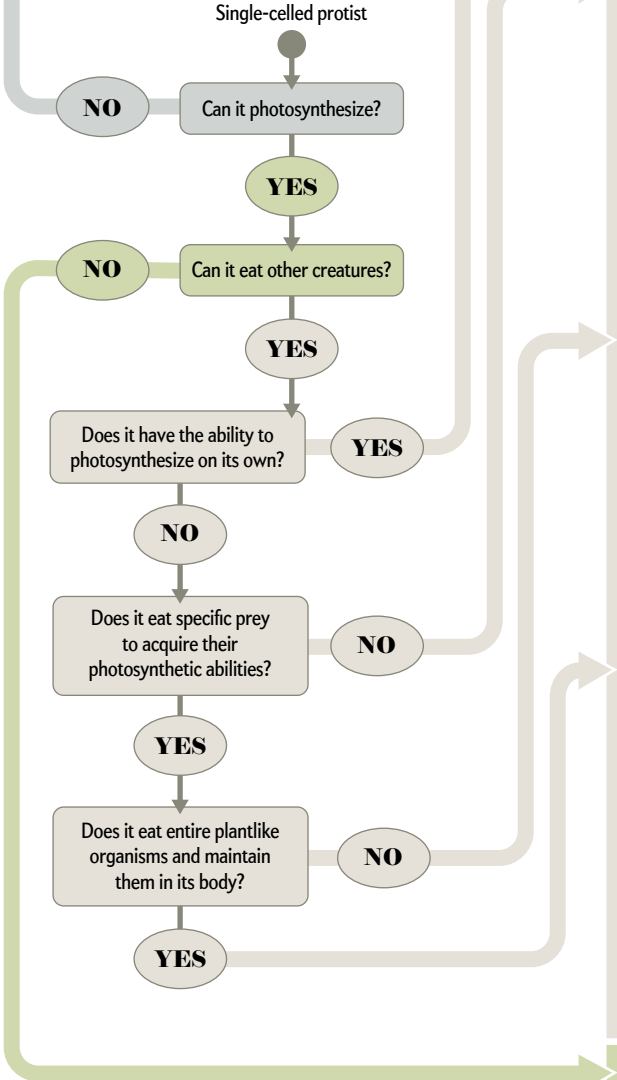
The nonconstitutive mixotrophs can be subdivided into generalist and specialist species. The plastidic ciliates *Laboea* and *Strombidium,* for instance, are generalists that steal chloroplasts from many different types of plankton. Generalists cannot maintain these looted parts for more than a few days and must constantly attack new prey for replacements. They tend to be helpful mixotrophs, contributing nutrients to food webs that support fisheries, and thus play a key role in global food security.

Specialists, on the other hand, depend solely on a particular type of prey and seem to be better equipped by evolution to integrate stolen photosystems into their own physiology; they can

© 2018 Scientific American

A New Plankton Menagerie

Microplankton, single-celled marine organisms, are one of the most critical forms of life on our planet, sustaining our global food web. Scientists used to think they were like either plants or animals. New evidence indicates most microplankton really live as mixotrophs: they combine plantlike photosynthesis with animallike hunting and eating in a variety of strategies. Now, to classify a single-celled plankton (also called a protist), biologists ask a series of questions:



Microzooplankton

These creatures have no plantlike abilities to sustain themselves with photosynthesis and so must eat other plankton to survive.

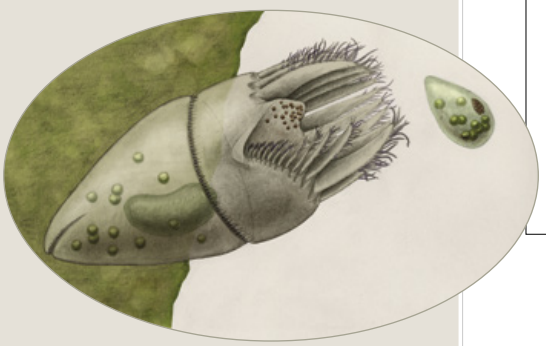
Mixotrophs

Constitutive Mixotroph

These predatory cells have physiology that also lets them use sunlight and nutrients to create energy. *Prymnesium parvum*, one such creature, is seen here attacking *Dunaliella tertiolecta*.

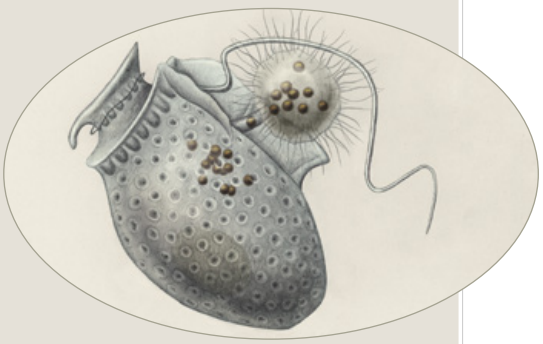
Generalist Nonconstitutive Mixotroph

These creatures use photosynthetic organelles taken from a variety of prey. Here *Strombidium oculatum* is stealing a body part from one victim, a member of the genus *Ulva*.



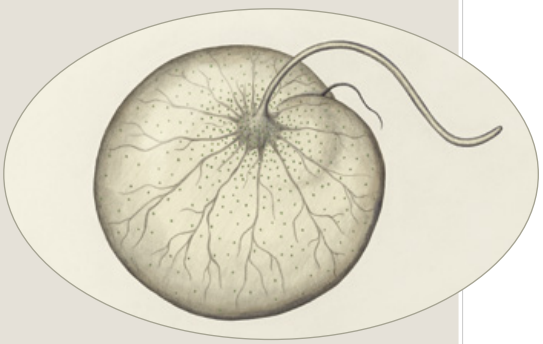
Plastidic Specialist Nonconstitutive Mixotroph

These predators steal parts from specific victims, killing them in the process. *Dinophysis acuminata* is seen sucking organelles from *Mesodinium rubrum*.



Endosymbiotic Specialist Nonconstitutive Mixotroph

Creatures in this group keep colonies of photosynthesizing prey within their body. Here *Noctiluca scintillans* is filled with *Pedinomonas noctilucae*.



Microphytoplankton

Organisms in this group act like land plants, surviving through photosynthesis.

SOURCE: "DEFINING PLANKTONIC PROTIST FUNCTIONAL GROUPS ON MECHANISMS FOR ENERGY AND NUTRIENT ACQUISITION: INCORPORATION OF DIVERSE MIXOTROPHIC STRATEGIES," BY ADITEE MITRA ET AL., IN *PROTIST*, VOL. 167, NO. 2, APRIL 2016

Illustration by Rebecca Gelernter

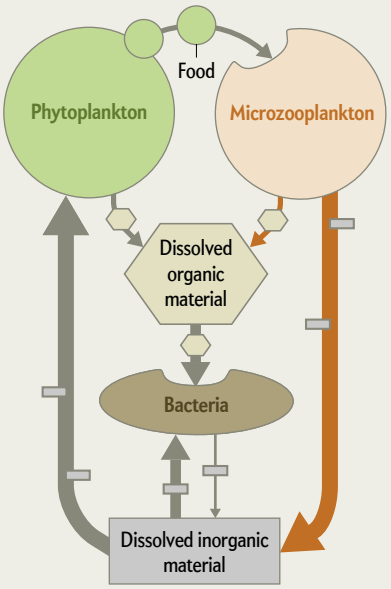
© 2018 Scientific American

The Mixotroph Effect

When scientists include mixotrophs among ocean plankton, rather than assuming plankton are either purely plantlike or animallike, their categorization results in big changes in the way that nutrients move through a food web and when populations of important microorganisms grow and shrink. Researchers learned this after they modeled the traditional two-type plankton ecology and compared it with a new model that included mixotrophs that plunder other plankton for photosynthesizing parts. The scientists also created a third model with mixotrophs that have innate photosynthesizing ability. The results of both mixotroph models more closely matched real-world observations than did the old, traditional model.

Traditional Scenario

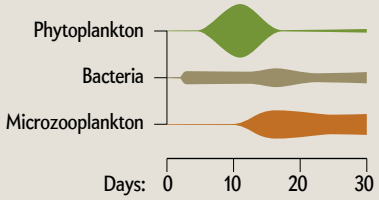
Phytoplankton use solar energy and inorganic material to make food, and microzooplankton eat the phytoplankton. Marine bacteria then break down released organic material—waste products—for reuse. It is a tight cycle that limits population sizes.



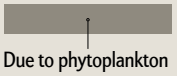
CLASSIC ONE-TWO SEQUENCE

Traditionally, over a 30-day cycle, phytoplankton (*green*) bloom first. Only then can predatory microzooplankton (*orange*) eat enough to increase their own group size. This hunger, of course, depletes the phytoplankton.

Changes in Plankton Populations

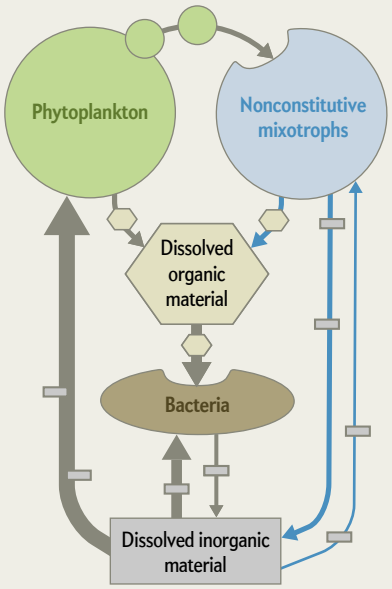


Carbon Dioxide Removed from Seawater
(Total: 30 grams of carbon per square meter)



Photosynthetic Animals

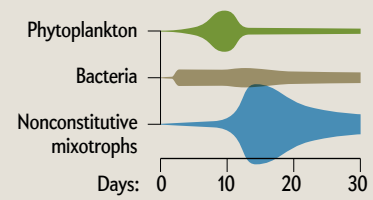
When animals that can also photosynthesize, the nonconstitutive mixotrophs, substitute for purely animallike microzooplankton, their hybrid abilities let them retain more nutrients. The thin blue arrow leading away from that group indicates less nutrient loss.



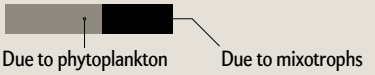
POPULATION BOOM

Nonconstitutive mixotroph populations (*blue*) can grow larger than traditional predators because they eat and also gain energy from stolen photosynthetic plankton parts. But growth tails off after they run out of parts to steal.

Changes in Plankton Populations

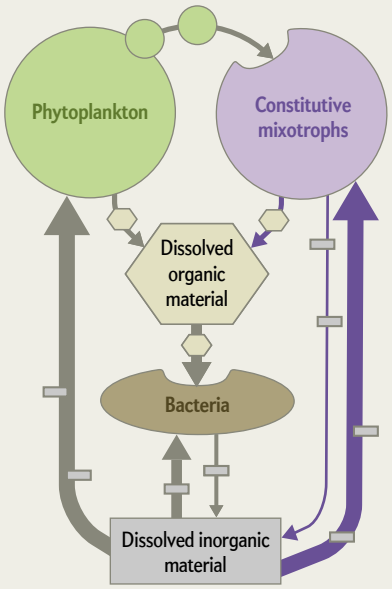


Carbon Dioxide Removed from Seawater
(Total: 30 grams of carbon per square meter)



Plants That Eat

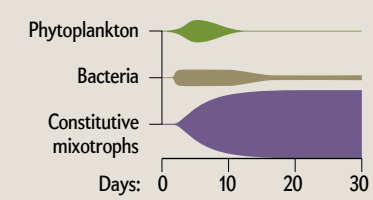
Replacing traditional microzooplankton with plants that can hunt, the constitutive mixotrophs, leads to a different population pattern. They can utilize lots of inorganic material, shown by the wide purple arrow, and act as predators. This sustains the group for an extended period.



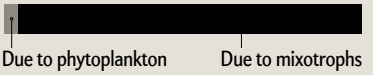
SELF-SUFFICIENT GROWTH

With innate photosynthetic abilities, constitutive mixotrophs (*purple*) need just a little help—extra food—from traditional phytoplankton to grow. Their population increases early and stays big through 30 days.

Changes in Plankton Populations



Carbon Dioxide Removed from Seawater
(Total: 65 grams of carbon per square meter)



SOURCE: "DEFINING PLANKTONIC PROTIST FUNCTIONAL GROUPS ON MECHANISMS FOR ENERGY AND NUTRIENT ACQUISITION: INCORPORATION OF DIVERSE MIXOTROPHIC STRATEGIES," BY ADITEE MITRA ET AL., IN *PROTIST*, VOL. 167, NO. 2; APRIL 2016

Graphic by Jen Christiansen

© 2018 Scientific American

maintain their pilfered assets for weeks to months. Some, such as *Dinophysis,* can be harmful to humans. Shellfish exposed to *Dinophysis* can give people life-threatening food poisoning, and large blooms, such as some that have occurred in the Gulf of Mexico, have closed oyster farms.

Some specialists can be split off into yet another group with a remarkable kind of behavior. They do not just steal body parts, like *Mesodinium* does, but instead take in and enslave entire colonies of photosynthetic prey. The colonies live and proliferate within the host, feasting on nutrients and enjoying protection from outside predators. These planktonic greenhouses include microorganisms called foraminiferans and radiolarians and are found throughout the oceans. Over hundreds of millions of years, foraminiferans have played a vital role in the global biological carbon pump, taking in large amounts of the element, sequestering it when they sink to the seafloor and later releasing it in smaller amounts as they decompose. Analyzing ancient layered deposits of foraminiferans has helped us to assemble records of past climate changes and to link them to mass extinction events. Not all these floating greenhouses are harmless, however. One type, the green *Noctiluca* species, is capable of causing harmful algal blooms in polluted coastal waters.

GLOBAL REGULATORS

FROM PLANTS that eat to animals that photosynthesize and from tiny two-micron organisms to relatively large one-millimeter plankton, mixotrophs span a wide range of ocean life. Why does this matter? Because it turns out that small organisms can have a variety of large impacts.

There is, for instance, an enormous nutrient-scarce area covering thousands of square kilometers in the middle of the central Atlantic Ocean. Scientists used to think that phytoplankton competed with marine bacteria for dissolved inorganic nutrients such as iron and phosphates in this area, which did not leave much to go around. But Mikhail Zubkov, a microbial biogeochemist at the National Oceanography Center in England, and his colleagues found a sizable population of constitutive mixotrophs—the ones that do their own photosynthesis—in this area when they sampled the waters during research cruises.

From these observations, Team Mixotroph developed two food web simulators. One was based on the traditional model of plant-bacteria competition, and the other added the mixotrophs. The team found that the mixotroph simulation was the one that best matched the nutrient amounts and cycles that Zubkov had observed. Instead of competing with plantlike phytoplankton, the bacteria grew using sugars and other nutrients that leaked out of the mixotrophs. Then the mixotrophs turned around and ate the bacteria, which gave them more phosphates and iron than they could pull from the seawater. And the model only fit observations if the mixotrophs were constitutive.

There was another key effect. With mixotrophs, and not traditional phytoplankton, the levels of carbon fixation—carbon dioxide taken out of seawater by the organisms—were significantly increased. The finding suggests that if these hybrids were not around, global amounts of carbon, which contribute to planetary warming, might be even higher.

Mixotrophs are especially important in coastal seas, where their effects on fisheries can be profound. In 2017, using a North Sea plankton model with mixotrophs of varying types, we found that when small mixotroph species eat marine bacteria, their populations grow larger, and as a result, they can outcompete other plankton that tend to form algal blooms. These scumlike blooms are not toxic, but they do block sunlight, putting a big crimp in a nutrient cycle that feeds tiny fish larvae and helps them to grow. Fewer blooms means more fish.

For fish health, it is also important that mixotrophs, as our observations have shown, turn out to be the dominant plankton form during summers. Pure plantlike plankton grow in the spring but then decline, so delicate fish larvae cannot depend on them. But mixotrophs are still around, and they are good, rich food that sustains the fish during this period.

A DUAL FUTURE

MIXOTROPHS ARE AT THE CENTER of so much in marine science, be it climate change and fisheries projections, reconstructing ancient time lines of carbon cycling or predicting destructive algal blooms. The challenge now is to use both real-time observations and our models to establish what the different mixotroph groups are doing in different locations during different seasons. This is important because as our climate changes, we need to know which environmental conditions would lead to a bloom of the toxic *Karlodinium,* or the ecologically damaging, green *Noctiluca,* or the fisheries-sustaining plastidic ciliates. We have recently completed the first steps toward this goal, mapping the presence of different mixotroph groups across the global seas. We next need to measure population sizes during varying seasons because changing light and temperature drastically affect their growth and proliferation.

There are still marine scientists who point out these conclusions rest on our simulations as much as they rely on real-world observations, and that is a valid critique. It is why we need to get more scientists to examine mixotroph activity beyond the lab, out on the high seas.

In 2017 I applied to the European Commission for a grant to train scientists to do this work. In stark contrast to my fellowship application 10 years earlier, this grant was funded, with glowing comments from scientific reviewers. Our expanding Team Mixotroph will be able to bring the next generation of marine researchers up to speed. Together we hope to figure out the many ways in which the oceans' perfect beasts may control our imperfect world. SA

Aditee Mitra specializes in mixotroph and zooplankton research and is a plankton systems dynamics modeler. She is a senior lecturer in bioscience at Swansea University in Wales and a visiting professor at the Free University of Brussels. For the latest mixotroph news, visit www.mixotroph.org

MORE TO EXPLORE

Building the "Perfect Beast": Modelling Mixotrophic Plankton. Kevin J. Flynn and Aditee Mitra in *Journal of Plankton Research,* Vol. 31, No. 9, pages 965–992; September 1, 2009.

Defining Planktonic Protist Functional Groups on Mechanisms for Energy and Nutrient Acquisition: Incorporation of Diverse Mixotrophic Strategies. Aditee Mitra et al. in *Protist,* Vol. 167, No. 2, pages 106–120; April 2016.

Mixotrophy in the Marine Plankton. Diane K. Stoeker et al. in *Annual Review of Marine Science,* Vol. 9, pages 311–335; January 2017.

scientificamerican.com/magazine/sa

© 2018 Scientific American

METAL WORLD: Asteroid Psyche, shown in an artist's conception, may be the iron-nickel core of an early planet precursor.

Our neighborhood of planets was not created slowly, as scientists once thought, but in a speedy blur of high-energy crashes, destruction and rebuilding

By Linda T. Elkins-Tanton

© 2018 Scientific American

SOLAR SYSTEM SMASHUP

© 2018 Scientific American

I WAS WALKING OUT OF A CLASSROOM AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY, where students and I had been talking about the way planets form, when I was stopped by my colleague Ben Weiss. He studies magnetism in space rocks, and he was very excited. Weiss pulled me down the hall to his office to show me new data on one of these rocks, a meteorite called Allende. It was information that could change almost everything planetary geologists thought about the solar system. It was 2009, and that fall Weiss's research team had shown that Allende—which crashed into Earth in a huge fireball in Mexico in 1969 and contained some of the oldest known material in our system—harbored signs of an ancient magnetic field in its rock. The discovery was a surprise. This kind of field, astronomers thought, was made only by a magnetic dynamo of intensely hot, flowing liquid metal inside a planet, the way Earth's magnetic field is produced by liquid iron spinning in the planet's core.

But Allende was supposed to be a fragment from a parent planetesimal—an early, nascent planet—that had been only slightly warm. Scientists presumed it had never gotten hot enough to melt the metal it contained. How, then, Weiss wondered, could this ancient piece of our solar system have become hot enough to create a magnetic dynamo?

My students had just been pelting me with questions about planetary evolution, challenging me to rethink some textbook wisdom, so I happened to have the bare bones of a new idea that might help answer Weiss's question. I went over to his whiteboard and began sketching it out.

Planetesimals had long been known to contain short-lived, unstable aluminum atoms radiating excess nuclear energy. This radioactive isotope is called ^{26}Al, and when it decayed, that excess energy could have heated planetesimals. Conceivably, the heat from ^{26}Al in Allende's parent body could have climbed so high that the object actually melted from the inside out. Metal within the body would have separated from silicate minerals in the rest of the rock and formed a liquid core that started to spin as the space rock rotated, creating a magnetic dynamo. Meanwhile the outside of the planetesimal would have been chilled by the cold of space, and cold rock and dust from our solar system's primitive disk would have kept adding to this unmelted rind.

This idea that early building blocks of the solar system contained so much energy was not the story I had learned in high school. Textbooks often still say that the solar system formed in a quiet, stately manner. That creation, 4.567 billion years ago, was thought to be an orderly, civilized process, like a minuet: the gas and dust of a molecular cloud spun down into a disk around a growing young star, and the gas and dust coalesced into many little boulders, each one of which gradually grew to tens to hundreds of kilometers in diameter. These planetesimals themselves then rammed together to form larger bodies, each perhaps the size of Mars, called planetary embryos. Only after that did the temperature in this space nursery pick up. These embryos,

IN BRIEF

Slow and steady growth, from small particles to large planets, was the way scientists envisioned the evolution of the solar system even within the past decade.

Recent analyses of meteorites, remnants of the system's birth, imply instead that the system came from chaotic collisions, melting and reconstruction.

Scientists plan to send a spacecraft to Psyche, a strange all-metal asteroid that may be a naked chunk of an early planet core, to verify this idea.

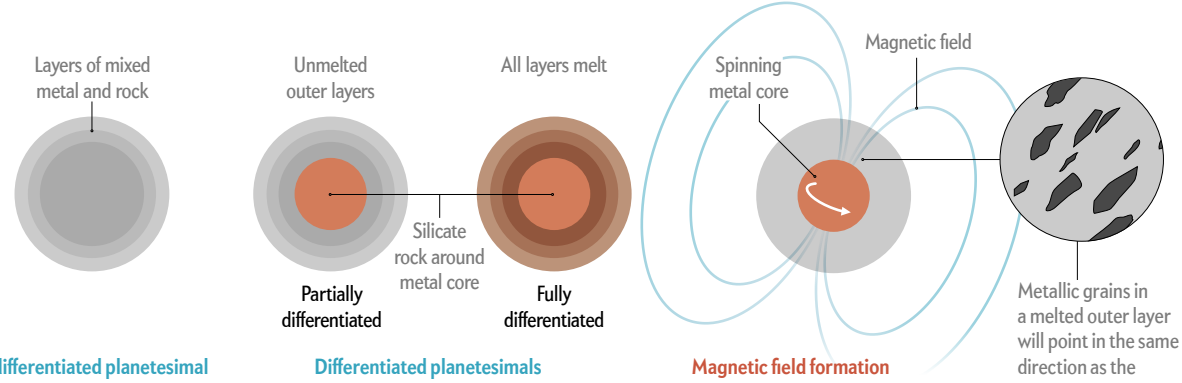
PRECEDING PAGES: COURTESY OF ASU AND PETER RUBIN *Iron Rooster Studios*

© 2018 Scientific American

Solar System Building Blocks

Planets like Earth and its neighbors got started when the cloud of gas and dust around our sun generated lumps known as planetesimals. These bodies were once thought to be quiet planetary seeds. But researchers now believe they crashed around and jump-started high-energy processes that gave rise to true planets.

Inside a Planetesimal


Layers of mixed metal and rock
Unmelted outer layers
All layers melt
Silicate rock around metal core
Partially differentiated
Fully differentiated
Spinning metal core
Magnetic field
Metallic grains in a melted outer layer will point in the same direction as the magnetic field and retain that same orientation after the dynamo has stopped and the outer layer has cooled.

Undifferentiated planetesimal

These rocks were 10 to a few hundred kilometers in diameter and contained radioactive elements that heated their innards. They had an onionlike structure.

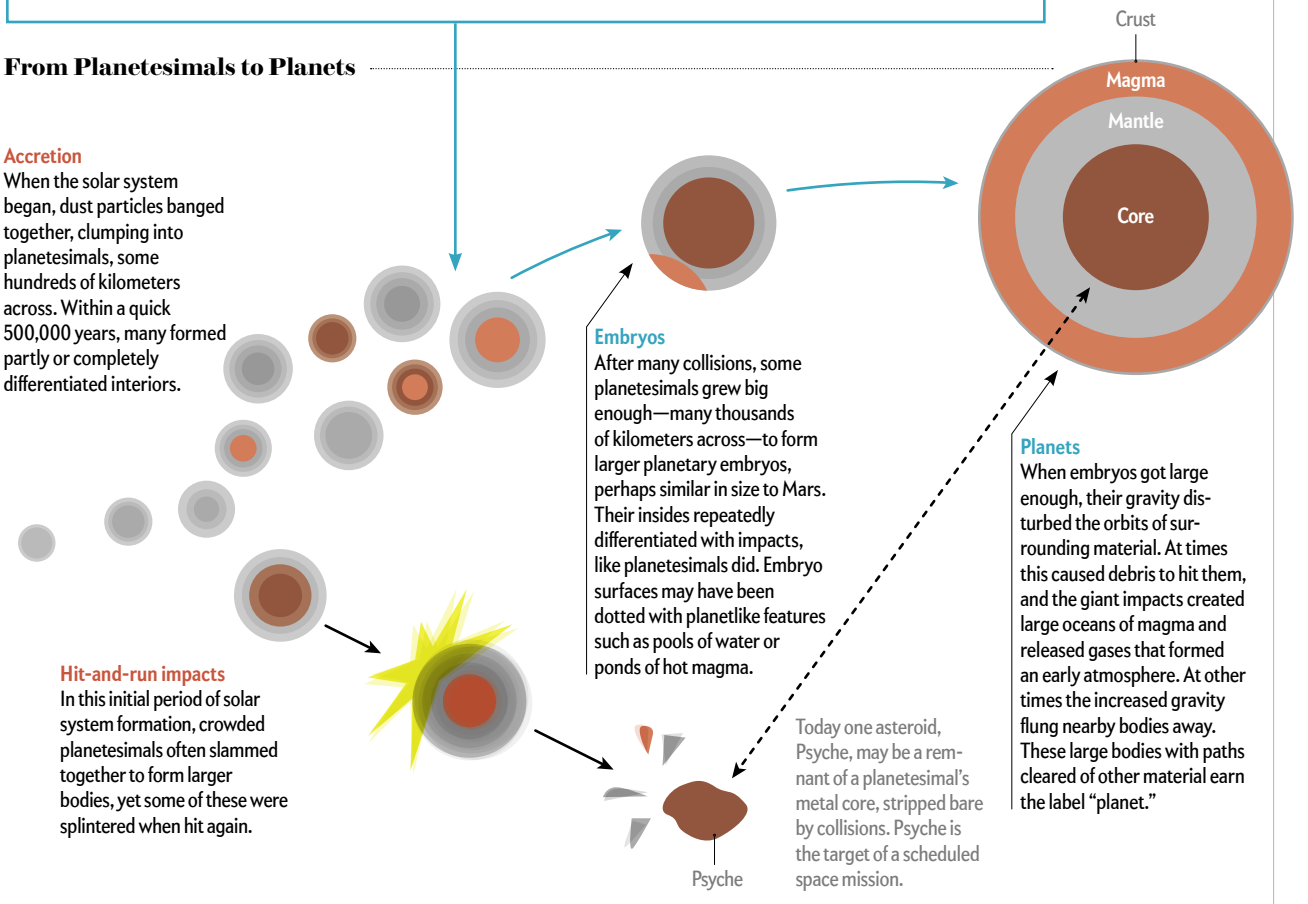
Differentiated planetesimals

Radioactive elements created more heat, and the body melted from the inside out. Denser iron-nickel metal sunk inward. Sometimes an outer rind was unaltered. All differentiated planetesimals had a liquid-metal core.

Magnetic field formation

Metal in motion can become a magnetic dynamo, which happened when planetesimals and their liquid metal cores started to spin. These cores generated magnetic fields pointing in particular directions, like Earth's field points north.

From Planetesimals to Planets



Accretion

When the solar system began, dust particles banged together, clumping into planetesimals, some hundreds of kilometers across. Within a quick 500,000 years, many formed partly or completely differentiated interiors.

Embryos

After many collisions, some planetesimals grew big enough—many thousands of kilometers across—to form larger planetary embryos, perhaps similar in size to Mars. Their insides repeatedly differentiated with impacts, like planetesimals did. Embryo surfaces may have been dotted with planetlike features such as pools of water or ponds of hot magma.

Hit-and-run impacts

In this initial period of solar system formation, crowded planetesimals often slammed together to form larger bodies, yet some of these were splintered when hit again.

Today one asteroid, Psyche, may be a remnant of a planetesimal's metal core, stripped bare by collisions. Psyche is the target of a scheduled space mission.

Planets

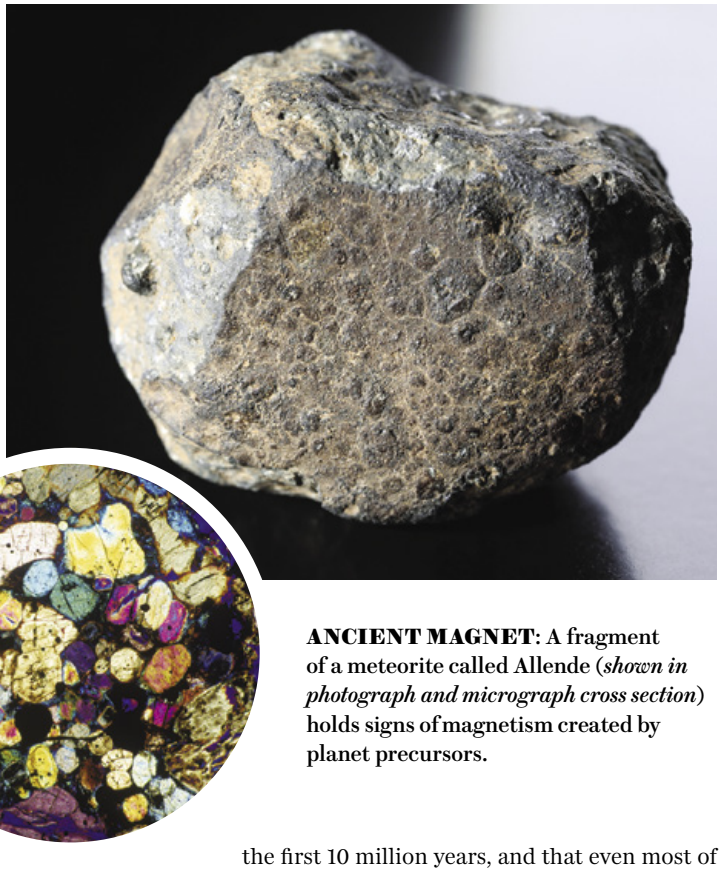
When embryos got large enough, their gravity disturbed the orbits of surrounding material. At times this caused debris to hit them, and the giant impacts created large oceans of magma and released gases that formed an early atmosphere. At other times the increased gravity flung nearby bodies away. These large bodies with paths cleared of other material earn the label "planet."

Illustration by Jen Christiansen

© 2018 Scientific American

which in growing had gained gravity strong enough to begin clearing their orbits of surrounding debris, then collided together and grew into planets. Eventually components in these planets separated into the familiar churning metal core and a silicate mantle—hot, wildly volcanic places, antithetical to life.

That's the old view. By the time Weiss and I began pondering Allende, some other data, too, indicated that the earliest solar system was really a place of rapid and violent change. Now that tame sequence of dust to boulders to planetesimals to embryos to planets is being replaced. Actual planetesimal formation, once surmised to take place over hundreds of millions of years, happened in just about three million years. If the age of our solar system, in human time, is now a day, this growing up happened during the very first minute. More energy in the early small components—energy like that provided by the aluminum heating and early collisions—means they did not have to wait and grow before developing different layers. Relatively minute planetesimals could host processes previously believed to be confined to planets, from melting and degassing to the creation of magnetic dynamos and volcanism.



ANCIENT MAGNET: A fragment of a meteorite called Allende (*shown in photograph and micrograph cross section*) holds signs of magnetism created by planet precursors.

And things in this system did not simply grow from small to large. Often large things blew apart into small again. If planet-size spheres formed in those early years, through collisions of these smaller high-energy bodies, then glancing, hit-and-run blows among planetesimals sometimes stripped or destroyed them instead. Their debris could hit other spheres, enlarging them to planet size. Planets could be built, torn asunder and rebuilt in just 10 million years or less.

THE DISAPPEARING DISK

PLANETARY SCIENTISTS such as me have pieced together this novel, three-ring circus vision of the active young solar system with great help from new tools for calculating the ages of meteorites as well as the ages of planetary dust clouds—similar to our primordial solar system—elsewhere in the cosmos.

Over the past 10 to 15 years scientists have developed instruments capable of measuring the elements that make up space rocks to parts per million or even less. Because we have a good idea of how long it takes radioactive elements to decay into their daughter elements, such measurements allow us to date when the planets and planetesimals that shed these fragments formed and changed. Teams all over the world—notably Alex Halliday, formerly at the Swiss Federal Institute of Technology Zurich and now at the University of Oxford; Thorsten Kleine of the University of Münster in Germany; Stein Jacobsen of Harvard University; Mary Horan and Rick Carlson, both at the Carnegie Institution for Science; and Richard Walker of the University of Maryland—began measuring collections of meteorites. The work made clear that planetesimals formed within the first few million years after the dusty disk began to cool, that many of our terrestrial planets could have formed within the first 10 million years, and that even most of Earth probably formed and differentiated into a core and mantle by a few tens of millions of years.

Other avenues of research yielded similar results. With ever improving telescopes, we can see young stars growing in other parts of our Milky Way galaxy, and in some cases, we can see the dust and gas disk from which a star and its planets grow. By estimating the ages of stars orbited by planets and comparing those measures with estimates for stars surrounded only by disks of dust and gas, researchers determined about 10 years ago that these disks last, on average, only three million years.

Planetesimals, therefore, have just three million years to grow on average. Any dust and gas not accreted onto rocks by then is lost into the star or out into space, and no more material for planet building is available. Considering that theorists used to think accretion took hundreds of millions of years, this is quite an acceleration!

More evidence for this timing comes from using the decay of radioactive elements like a clock, which ticks away at a steady rate as one element turns into another. The new instruments gave teams throughout Europe and the U.S. enough precision to measure these elements and thereby learn how long that clock had been ticking. Meteorites that have fallen to Earth contain those elements. Most are pieces of asteroids, which themselves are primitive remnants of planetesimals. (Some other meteorites are from the moon, some are from Mars and others are from bodies not yet identified.)

One radioactive isotope of the element hafnium prefers to stay in silicate minerals, such as those in Earth's mantle. But it decays to an isotope of tungsten, which combines readily with metals that make up planet cores. This decay happens on a reg-

DETLEV VAN RAVENSWAAY *Science Source (photograph)*; MICHAEL ABBEY *Science Source (micrograph cross section)*

© 2018 Scientific American

ular schedule: half the hafnium turns into tungsten in nine million years. This system gives us the timescale of metal-silicate (core-mantle) differentiation of planets and planetesimals: metal core formation scavenges tungsten from the silicate mantle and carries it into the growing core. Any hafnium stays happily in the mantle and keeps on decaying into tungsten, which then remains in the silicate mantle if no further core formation occurs. When scientists measure the ratios of hafnium to tungsten in meteorites, the amount of that tungsten isotope in them gives the time since core formation.

Such isotopic measurements of iron meteorites—many of which presumably hail from the metallic cores of planetesimals—show that their parent bodies formed within just 500,000 years of the first solids condensing from our protoplanetary disk of dust and gas. That time is less than 10 seconds into our putative 24-hour solar system. If iron meteorites are core fragments of planetesimals smashed apart by impacts, then planetesimals must have actually formed, melted and created iron cores within that tiny period.

WITH EXPERIMENTAL DATA FIRMLY SHOWING THAT THE SOLAR SYSTEM FORMED MUCH MORE QUICKLY THAN TEXTBOOK SCENARIOS INDICATED, SCIENTISTS HAD TO EXPLAIN HOW PLANETS CAME INTO BEING SO QUICKLY. THAT PUT THE BALL IN THE THEORISTS' COURT.

GROWTH SPURT

WITH EXPERIMENTAL DATA firmly showing that the solar system formed much more quickly than textbook scenarios indicated, scientists who study it now had to explain how planets came into being so quickly. That put the ball in the theorists' court. How can dust and pebbles just microns to centimeters in diameter, all orbiting around our young sun, clump together to form bodies up to 10 million times larger (100-kilometer planetesimals) in just 500,000 years?

The answer is not obvious. Standard physics says that small clumps of dust that collide can readily stick together through electromagnetic forces, much as static electricity creates clumps of household dust. The energy absorbed as a result of compression and loss of porosity during collisions also helps clumps stick together rather than bouncing off or breaking apart. As the clumps grow, though, they reach what is called the meter barrier. Before they reach a meter in diameter, these growing boulders become too large to stick by electromagnetic forces and too small still to stick through gravitational attraction. Impacts even at very low speeds cause these conglomerates to destruct rather than accrete. Yet we know that such materials must be able to grow from meter size to planetesimal size—the planet we stand on today is a testament to that. So some other processes must have been at work.

Several ideas for how growth beyond the meter barrier occurs have been proposed. Most hypotheses involve concentrating material in the protoplanetary disk through various types of turbulence that slam particles together. Such whirlpool-like forces might include phenomena called Kelvin-Helmholtz eddies, which develop between the gas and dust layer of the disk and could effectively crush regions of material together into larger bodies. Much of that work has been pioneered by Anders Johansen, now at Lund University in Sweden. Hal Levison of the Southwest Research Institute and Johansen have separately worked on a newer model, called pebble accretion. His calculations indicate that even the smallest pieces of dust and clumps can be gravitationally diverted over several orbits to add to a growing planetesimal and can do so fast enough to build planetesimals early in the life of the solar system.

MELTING IN A FREEZER

NO KIND OF CRUSHING, however, could have caused the planetesimals to differentiate into mantles and cores. If planetesimals were first formed of primordial disk material in which metals and silicates were intimately mixed, then only high temperatures and at least partial internal melting would allow the metal to sink to the interior and form a core. Calculations show that slamming these relatively small bodies together would not provide enough energy to melt them. Consequently, researchers were left wondering where that melting energy—within the vast freezer of chilly space—came from.

This is where ideas about the radioactive aluminum come in. Each time one of these atoms decays, it gives off a little burst of heat. These tiny quanta of heat could have added up to a powerful energy source in the early solar system. Because aluminum is one of the six most common elements in stony material (the others are silicon, magnesium, iron, oxygen and calcium), ^{26}Al, which has a half-life of approximately 700,000 years, could easily have heated at least some planetesimals to a melting temperature.

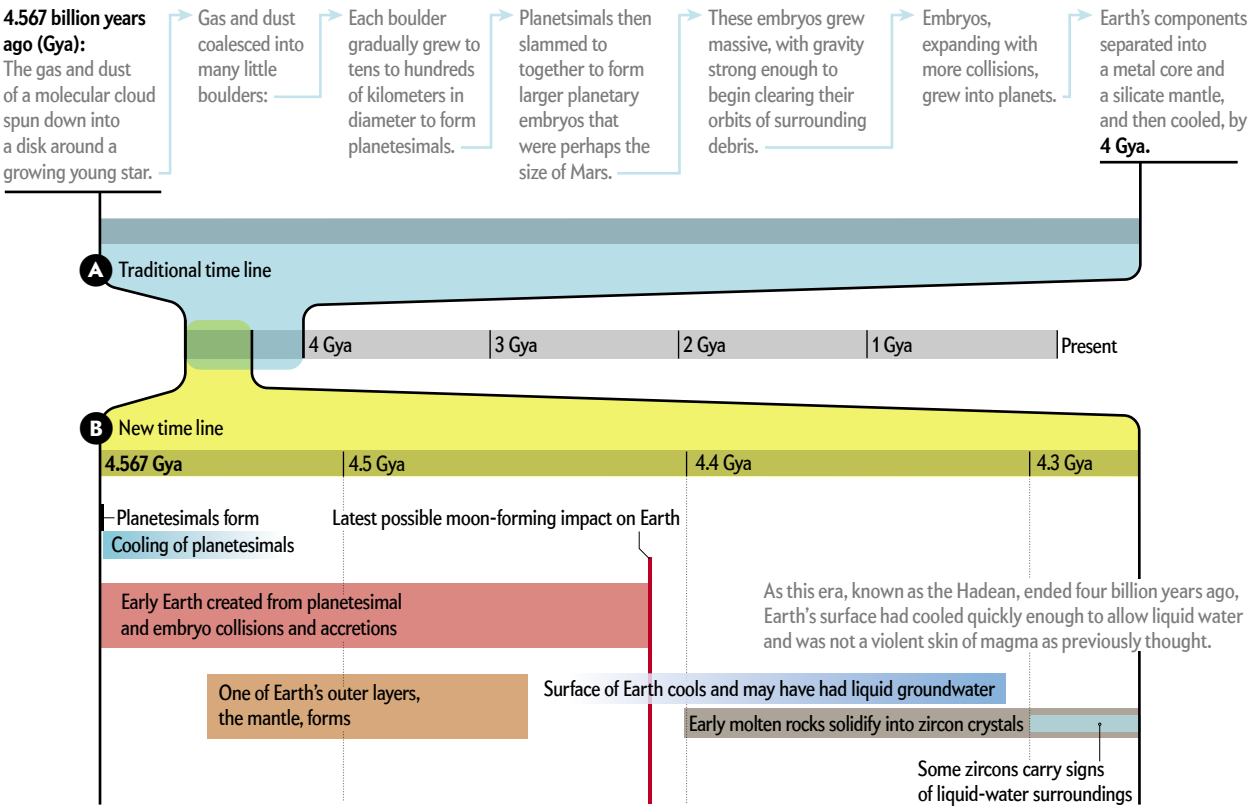
But what would have kept them from melting completely, considering that the new observations strongly indicate that some of these planetary seeds had unmelted rinds on the outside? Size is part of the answer. In the case of Allende's parent planetesimal, to reach melting, the rocky body would have needed to be massive enough so that its interior produced heat faster than its exterior radiated it away. A larger body can get hotter inside than a smaller body can because it has a greater heat-producing volume compared with its heat-releasing surface. But the short half-life of ^{26}Al means this growth had to be quick. To retain enough heat for the melting pattern we envisioned, Allende's parent planetesimal would have grown to 10 kilometers or more in radius within about two million years of the first solids in the solar system (the equivalent of the first 37 seconds of our 24-hour solar system)—and we think it may have grown to as much as 200 kilometers in radius.

It had been thought that planetesimals either melted entirely or remained primitive. But Weiss and I were suggesting a hybrid, where the most primitive material in the solar system encased a planetesimal that had melted inside—an outer rind

© 2018 Scientific American

Speeding Up Construction

Scientists used to think the formation of planets in our solar system was a slow, steady process that stretched over 500 million years, beginning at about 4.5 billion years ago **A**. New evidence from meteorites that fell to Earth, decay rates of radioactive elements and observations of dust around stars indicate our system—including a nascent Earth—got started much more quickly **B**, within three million years, in a high-energy blur of protoplanetary crashes and rebuilding.


4.567 billion years ago (Gya): The gas and dust of a molecular cloud spun down into a disk around a growing young star.
Gas and dust coalesced into many little boulders:
Each boulder gradually grew to tens to hundreds of kilometers in diameter to form planetesimals.
Planetesimals then slammed to together to form larger planetary embryos that were perhaps the size of Mars.
These embryos grew massive, with gravity strong enough to begin clearing their orbits of surrounding debris.
Embryos, expanding with more collisions, grew into planets.
Earth's components separated into a metal core and a silicate mantle, and then cooled, by 4 Gya.
A Traditional time line
4 Gya
3 Gya
2 Gya
1 Gya
Present
B New time line
4.567 Gya
4.5 Gya
4.4 Gya
4.3 Gya
Planetesimals form
Cooling of planetesimals
Latest possible moon-forming impact on Earth
Early Earth created from planetesimal and embryo collisions and accretions
As this era, known as the Hadean, ended four billion years ago, Earth's surface had cooled quickly enough to allow liquid water and was not a violent skin of magma as previously thought.
One of Earth's outer layers, the mantle, forms
Surface of Earth cools and may have had liquid groundwater
Early molten rocks solidify into zircon crystals
Some zircons carry signs of liquid-water surroundings

as well as a melted core. This made sense because the Allende meteorite—with its record of a magnetic field caused by a heated interior—consists of unheated, primitive material alone. The only place it could come from would be a cool outer rind. Allende's parent planetesimal retained this primitive unmelted surface rind because it was chilled by the coldness of space and because dust in the cool protoplanetary disk continued to attach to it over time. Unchanged by heating, the rind was able to maintain a record of the magnetic field produced by the inner part of the planetesimal's structure, the melted core and its magnetic dynamo.

We were not the first to have thought of partial differentiation. Geologist John Wood drew a similar structure by hand in his Ph.D. thesis at M.I.T. in 1958—but no one had ever been heretical enough to say that the poster child for primitive unmelted meteorites, Allende, could have come into existence in this way or that the process was common, indeed formative, at the beginning of our solar system.

Now it does appear common. Researchers have found that at least four other meteorite parent bodies hosted magnetic core dynamos. At the same time, other possible sources of the magnetization have been ruled out: Allende and its fellows were not magnetized by a field produced in the sun, nor by the dusty disk itself, nor by transient plumes around impacts. If the earliest solar system was truly populated by hundreds or even thousands of differentiated planetesimals, zipping around and generating intense heat and magnetic dynamos like tiny Earths, it implied the entire infant system contained a great deal more heat than geologists had ever thought.

A CROWDED FIELD

OTHER IDEAS have also weakened the traditional notion of linear planet growth from small to large. For years, for the sake of simplicity and numerical tractability, all simulations of planet formation assumed that in every collision of planetesimals, all the material from both colliders combined to form a new, bigger body. This merging occurred even as planetesimals were forming from dust. But new understanding and new approach-

Graphic by Jen Christiansen

© 2018 Scientific American

es to modeling the collision process have been pioneered by Erik Asphaug of Arizona State University. Some collisions are constructive, Asphaug found, and produce larger bodies. But other crashes can be destructive, with the collider stripping material from the target body and continuing on its way to do more damage elsewhere.

Only at about 10 million years of age did bodies really get bigger and stay big. What gave them enough stability to survive? Again, the answer seems to be size. As planetesimals collided and formed larger planetary embryos, their mass and therefore gravity became larger. The gravity was great enough that whenever their orbits brought them near another object, that object was either drawn in and accreted through the pull of gravity or was flung away as its orbit was changed. Thus, these growing planets began to clear out their orbits, which is one of the criteria for being properly called a planet. Smaller bodies had fewer and fewer refuges where their orbits could remain stable and unperturbed by growing planets; the asteroid belt became one of the safe havens left to them.

WE WANTED TO EXPLORE A PLACE THAT COULD PROVE OR OVERTURN HYPOTHESES. THE BEST DESTINATION, WE HAD DECIDED, WAS A WORLD MADE OUT OF METAL: THE METAL ASTEROID PSYCHE. THERE IS NOTHING QUITE LIKE THIS PLACE, AT LEAST NOTHING CLOSE ENOUGH TO REACH IN A REASONABLE AMOUNT OF TIME.

MISSION TO A METAL WORLD

WEISS, ASPHAUG, I and others would like to learn how our own planet's structure and composition came into being out of this energetic, often chaotic environment. But we cannot, despite the science fiction of popular movies or Jules Verne novels, get a good look at Earth's core. It is too deep and has pressures too high to sample directly.

But perhaps one particular asteroid, a remnant of an ancient planetesimal, might be a decent stand-in. About six years ago several of my colleagues and I began to design a space mission to explore this possibility. We gathered in the Left Field mission-formulation room at NASA's Jet Propulsion Laboratory. This room is designed for creativity. It has shelves full of drawing and building materials such as cardboard, wheels, wire, Legos, paper, markers and foam. It is a good room to think about something completely new, and that is what we were trying to do. We wanted to explore a place that could prove or overturn hypotheses. The best destination, we had decided, was a world made out of metal: the metal asteroid Psyche.

There is nothing quite like this place, at least nothing close enough to reach in a reasonable amount of time. Psyche is one of the largest asteroids, about 200 kilometers in diameter, and is located between Mars and Jupiter. All the physical measurements we have—from radio telescopes bouncing waves off the body—indicate that it consists almost solely of iron and nickel. Psyche looks like it is a stripped-naked planetesimal core, a last remnant of the hit-and-run collisions that disrupted bodies in the early solar system. The orientation of particles in Psyche, like tiny magnetic compass needles, might tell us whether it had a magnetic dynamo. There might also be some remnants of its rocky exterior that tell us what the deep mantle of planetesimals looked like. If there were surface impacts on the naked metal, the splashes might have produced sharp metal cliffs that froze before they could fall back to the surface.

Each of us in the mission room brought a particular set of skills to the table: Weiss, his specialty of measuring magnetic fields in meteorites; William F. Bottke, the dynamics of orbiting bodies; Asphaug, the effects of collisions; Bruce Bills, calculating the gravity field of a body; Daniel Wenkert, managing data and operations. Damon Landau calculated trajectories; he is an interplanetary travel agent. John Brophy organized our deliberations, and I brought my knowledge of compositions, melting, solidification and differentiation processes.

The energy in the room rose to a heady peak. There was no e-mail checking, and there were no conversational pauses. We were united in our excitement over true exploration: humankind has never visited a metal body, and we do not even know what it will look like.

Now, years later, those days of playing with foam and Legos have given way to an organized push by a team of about 150 people and growing. We proposed a spacecraft about the size of a singles tennis court, powered by solar cells and ion thrusters and carrying a magnetometer to detect magnetic fields, a gamma-ray spectrometer to identify elements and two cameras. In January 2017 we received the good news that NASA had selected the Psyche mission for flight. We will launch the spacecraft in 2022 and project it will arrive at this strange remnant of planet building in 2026. The world of metal awaits. SA

Linda T. Elkins-Tanton, a planetary geologist specializing in the evolution of terrestrial planets, is director of the School of Earth and Space Exploration at Arizona State University.

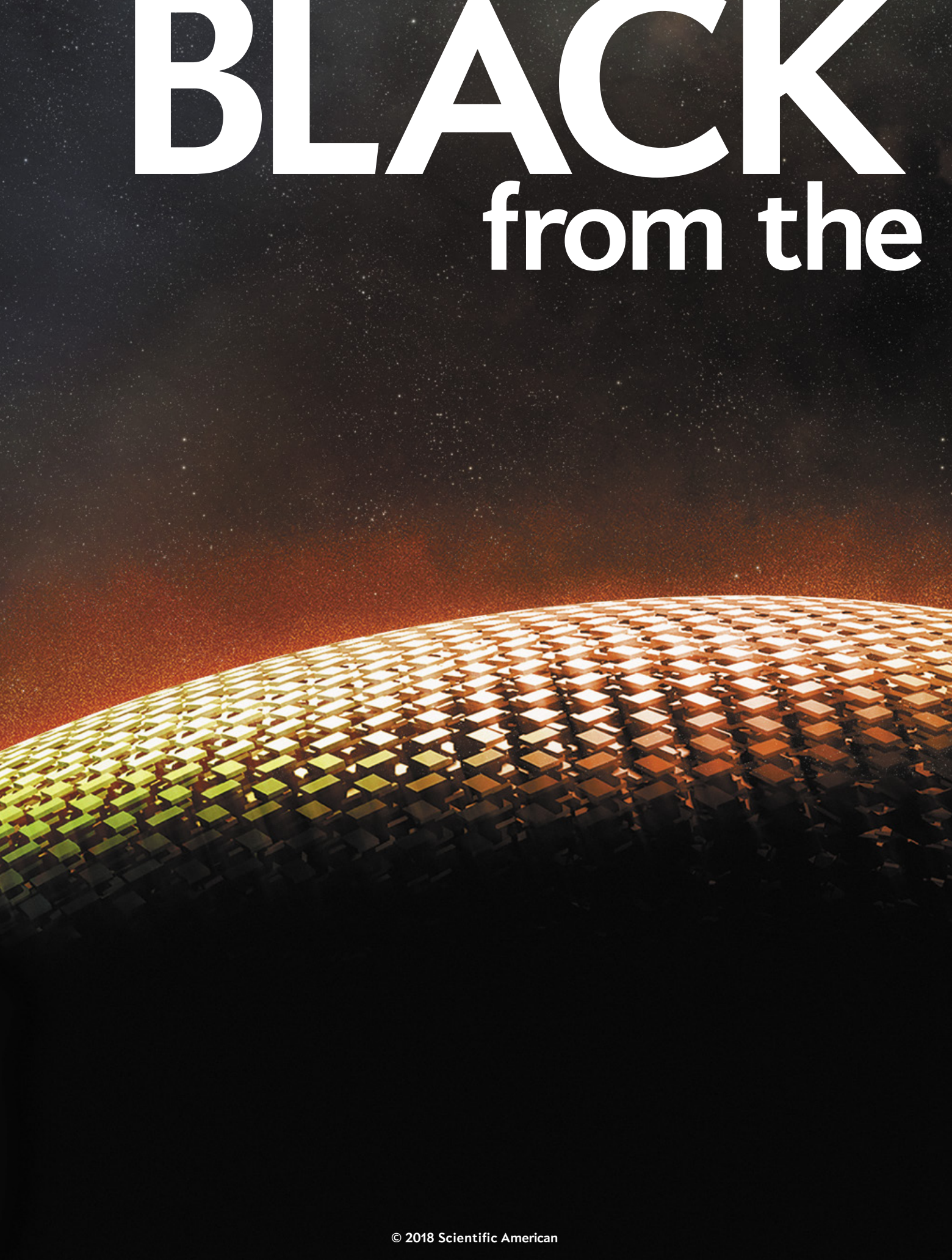
MORE TO EXPLORE

Iron Meteorites as Remnants of Planetesimals Formed in the Terrestrial Planet Region. William F. Bottke in *Nature,* Vol. 439, pages 821–824; February 16, 2006.
Asteroids, Meteorites, and Comets. Revised edition. L. T. Elkins-Tanton. Facts On File, 2010.
The Earth and the Moon. Revised edition. Linda T. Elkins-Tanton. Facts On File, 2010.
Jupiter and Saturn. Revised edition. Linda T. Elkins-Tanton. Facts On File, 2010.
Mars. Revised edition. Linda T. Elkins-Tanton. Facts On File, 2010.
The Sun, Mercury, and Venus. Revised edition. Linda T. Elkins-Tanton. Facts On File, 2010.
Uranus, Neptune, Pluto, and the Outer Solar System. Revised edition. Linda T. Elkins-Tanton. Facts On File, 2010.
Vision and Voyages for Planetary Science in the Decade 2013–2022. Committee on the Planetary Science Decadal Survey. National Academies Press, 2011.
Mercury and Other Iron-Rich Planetary Bodies as Relics of Inefficient Accretion. E. Asphaug and A. Reufer in *Nature Geoscience,* Vol. 7, No. 8, pages 564–568; August 2014.
NASA Selects Two Missions to Explore the Early Solar System. Press release. January 4, 2017. www.nasa.gov/press-release/nasa-selects-two-missions-to-explore-the-early-solar-system

scientificamerican.com/magazine/sa

© 2018 Scientific American

BLACK from the



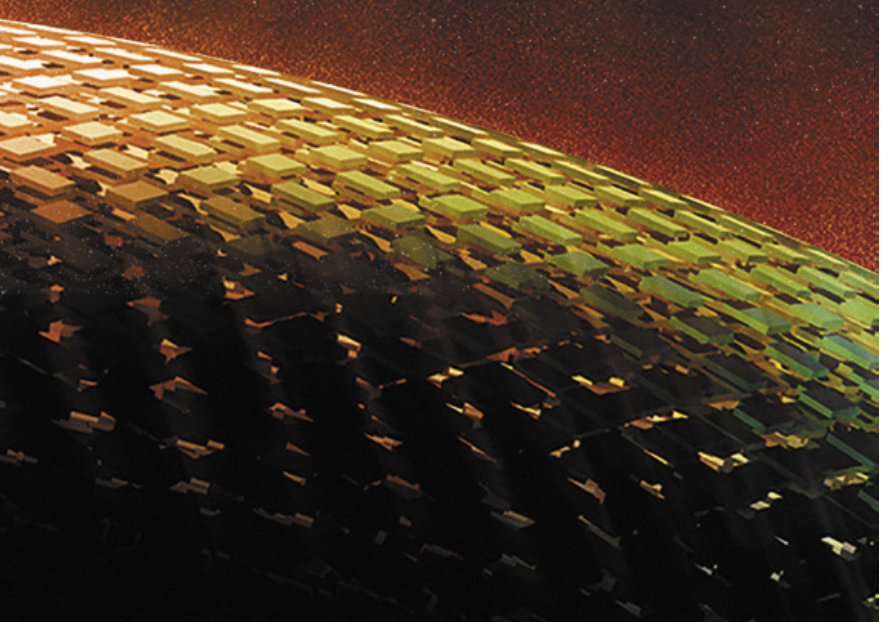
© 2018 Scientific American

HOLES Beginning of Time

A hidden population of black holes born less than one second after the big bang could solve the mystery of dark matter

By Juan García-Bellido and Sébastien Clesse

Illustration by Kenn Brown, Mondolithic Studios



IN BRIEF

The nature of dark matter—the invisible material that holds galaxies together by its gravity—is a deep cosmic enigma. **Many researchers suspect** dark matter is made of weakly interacting massive particles and have been seeking them in experiments. But to date, no such "WIMPs" have been found. **"Primordial" black holes** that may have formed shortly after the big bang are an alternative candidate for dark matter. Yet these, too, have so far eluded detection. **More evidence** for primordial black holes may emerge in new data from gravitational-wave detectors and other observatories. If confirmed to exist, these objects could solve the mystery of dark matter and several other cosmic conundrums.

More than a billion years ago two black holes in the distant universe spiraled around each other in a deathly dance until they merged. This spiraling collision was so violent that it shook the fabric of space-time, sending perturbations—gravitational waves—rippling outward through the cosmos at the speed of light. In September 2015, after traveling more than a billion light-years, those ripples washed over our planet, registering as a "chirp" in the sensors of the Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO).

This was the first direct detection of gravitational waves, and the observation confirmed Albert Einstein's century-old prediction of their existence. Yet the chirp revealed that each of the merger's progenitor black holes was 30 times heavier than the sun. That is, their masses were two to three times larger than ordinary black holes born from supernova explosions of massive stars. These black holes were so heavy, it is hard to explain how they formed from stars at all. Furthermore, even if two such black holes did independently form from the deaths of very massive stars, they would have then had to find each other and merge—an event with an exceedingly low probability of occurring within the current age of the universe. It is thus reasonable to suspect that these massive black holes formed via some other, more exotic pathway that might

© 2018 Scientific American

not involve stars at all. Beyond its detection of gravitational waves, it may be that LIGO has unveiled something even more extraordinary: black holes that predate the formation of the stars themselves.

Although such "primordial" black holes have never before been seen, some theoretical models suggest they could have formed in astronomical numbers from the hot, dense plasma that filled the cosmos less than one second after the big bang. This hidden population could solve several outstanding mysteries in modern cosmology. In particular, primordial black holes could constitute some, if not all, of dark matter—the invisible 85 percent of the matter in the universe that acts as gravitational glue to hold galaxies and galaxy clusters together. Further studies with LIGO and other facilities will soon test these ideas, potentially unleashing a new revolution in our understanding of the cosmos.

THE FALL OF MACHOS, THE RISE OF WIMPS

BLACK HOLES would initially seem to be ideal candidates for dark matter because they emit no light. Indeed, along with other dark objects such as planets and brown dwarfs, they make up one long-proposed solution to the dark matter problem: MACHOs, short for *ma*ssive *c*ompact *h*alo *o*bjects. Found both in spherical halos surrounding each galaxy and near each galaxy's luminous center, MACHOs would create the gravitational pull responsible for the otherwise anomalous motions of stars and gas that astronomers observe in the outskirts of galaxies. Simply put, galaxies seem to be rotating too fast to be held together by the visible mass in stars that we observe. Dark matter provides the extra pull to prevent spinning galaxies from flinging off their stars.

If MACHOs make up most of the universe's dark matter, they must also account for other observations. Whatever dark matter is, it shapes the universe's largest structures, determining the origin and growth of galaxies as well as clusters and superclusters of galaxies. These objects coalesce from the gravitational collapse of clumps of gas inside dark matter halos. Cosmologists have precisely mapped the spatial distribution of these clumps through deep and wide galaxy surveys and correlated them with tiny temperature fluctuations present in the cosmic microwave background (CMB), the big bang's all-sky afterglow. The diffuse mass of dark matter in large galaxies and clusters also bends space to distort the light from far distant background objects—a phenomenon known as gravitational lensing.

The MACHO hypothesis, however, fell from favor a decade ago when MACHOs did not turn up in tentative, indirect searches for their existence. Most notably, astronomers looked for them via microlensing, a variety of gravitational lensing in which a black hole, a brown dwarf or even a planet passes in front of a background star and temporarily magnifies the star's light. Several multiyear microlensing surveys of millions of stars in the Large and Small Magellanic Clouds, the main satellite galaxies of the Milky Way, found no evidence that MACHOs made up the entirety of our galactic halo. These results were conclusive enough to rule out MACHOs up to around 10 solar masses as the primary constituent of dark matter. As these surveys took place, theorists built the case for an alternative hypothesis—WIMPs, or weakly interacting massive particles.

WIMPs are predicted by certain extensions of the Standard Model of particle physics, but they remain at least as elusive as MACHOs. To date, no evidence of their existence has been found despite decades of searches using particle accelerators, underground detectors and space telescopes. As null results piled up in the search for WIMPs, some researchers began reconsidering the MACHO hypothesis, focusing particularly on primordial black holes. But what process could have seeded these strange objects throughout the observable universe, and how could they have eluded detection for so long?

BLACK HOLES FROM THE BIG BANG

PHYSICISTS BERNARD CARR and the late Stephen Hawking proposed the idea of primordial black holes in the 1970s. They considered black holes with masses smaller than that of a mountain. Such minuscule black holes would have already evaporated and vanished within the age of our nearly 14-billion-year-old universe, via a quantum-mechanical process discovered by Hawking and appropriately called Hawking radiation. But they also investigated the possibility that more massive, nonevaporating black holes could constitute the missing matter in galaxy clusters. The possibility that massive primordial black holes could actually be most or even all of the dark matter hinges on an idea known as cosmic inflation, first proposed by physicist Alan Guth in the early 1980s.

Inflation is a hypothetical phase of prodigious expansion immediately after the big bang. In 10^{-35} second, two points separated by less than an atomic radius would have become separated by four light-years, a distance comparable to that of the closest stars. Moreover, during inflation tiny quantum fluctuations are magnified to macroscopic scales by the rapid expansion, seeding the growing universe with underdense and overdense regions of matter and energy from which all cosmic structures subsequently emerge. As bizarre as it may seem, the theory of inflation is strongly supported by observations of such density fluctuations in the CMB.

In 1996 one of us (García-Bellido), together with Andrei Linde of Stanford University and David Wands of the University of Portsmouth in England, discovered a way for inflation to form sharp peaks in the spectrum of density fluctuations in the early universe [*see box on opposite page*]. That is, we showed how quantum fluctuations enormously magnified by inflation would naturally produce particularly dense regions that would collapse to form a population of black holes less than one second after inflation ends. Such black holes would then behave as dark matter and would dominate the matter content of the present-day universe. This model generated a population of black holes with the same mass, determined by the amount of energy within the collapsing region. Many other groups then started exploring these ideas within different models of inflation.

In 2015 the two of us (Clesse and García-Bellido) proposed a scenario, similar to that of 1996, in which these primordial fluctuations exhibit a broad peak in their energy densities and spatial sizes, giving rise to primordial black holes with a wide range of masses. A key consequence of this scenario is the fact that large density fluctuations collapse in close spatial proximity to one another, generating clusters of black holes of different masses—from one 100th to 10,000 times the mass of our sun. Within half a million years of the big bang, each growing, evolving cluster could contain millions of primordial black holes in a volume just hundreds of light-years across.

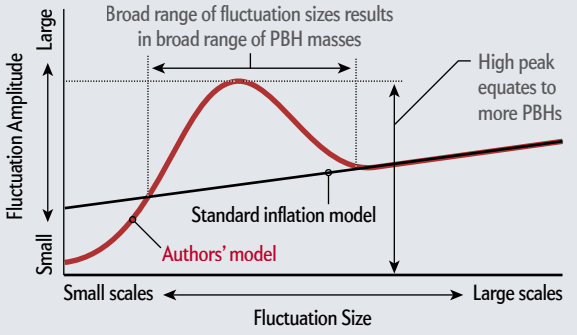
© 2018 Scientific American

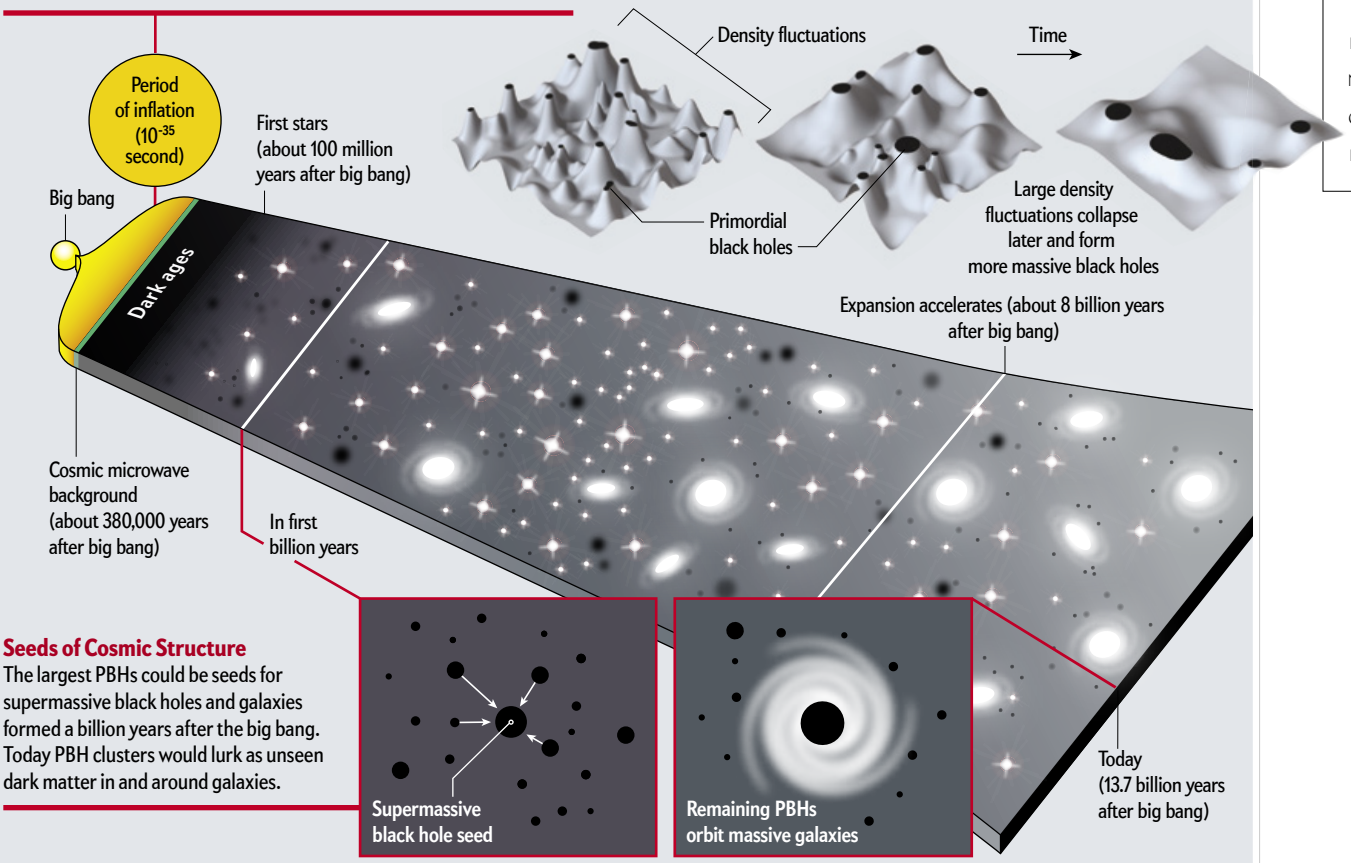
Black Holes Birthed by the Big Bang

The universe's first black holes may have been born in the earliest moments of cosmic time, when all was a seething, thick fog of fundamental particles. In the 1970s theorists realized that dense regions of that fog could collapse under their own gravity just a second after the big bang, forming so-called primordial black holes (PBHs) that would then shape the structure of the evolving, expanding universe. Emitting no light, PBHs would be a natural—albeit difficult to detect—candidate for dark matter.

Primordial Black Holes Form in Clusters

Inflation—a proposed acceleration to the universe's expansion less than a second after the big bang—would form PBHs by magnifying quantum fluctuations to immense scales. As inflation ended, these fluctuations would create density perturbations that then form PBHs. Larger, more powerful fluctuations would create more massive and numerous PBHs. The authors' inflationary model predicts a broad peak of magnified fluctuations and a range of density perturbations, producing PBHs in clusters, with each PBH ranging from one 100th to 10,000 times the mass of our sun. Half a million years after the big bang, a cluster might span hundreds of light-years and contain millions of PBHs. As the PBHs within such clusters merged together, scattered apart, and fed on ordinary gas and dust, they would guide the growth of galaxies and galactic clusters.




Density fluctuations
Time
Period of inflation (10^{-35} second)
First stars (about 100 million years after big bang)
Big bang
Dark ages
Primordial black holes
Large density fluctuations collapse later and form more massive black holes
Expansion accelerates (about 8 billion years after big bang)
Cosmic microwave background (about 380,000 years after big bang)
In first billion years
Today (13.7 billion years after big bang)

Seeds of Cosmic Structure

The largest PBHs could be seeds for supermassive black holes and galaxies formed a billion years after the big bang. Today PBH clusters would lurk as unseen dark matter in and around galaxies.

Such clusters of primordial black holes would be sufficiently dense to explain LIGO's mysterious black hole mergers, which one would not otherwise expect to occur with regularity. From time to time, the trajectories of two primordial black holes within a cluster can cross, so that both objects become gravitationally bound to each other. They would then spiral closer together for up to millions of years, radiating gravitational waves until they merged. In January 2015 we actually predicted that LIGO would detect gravitational waves from such massive mergers—waves identical to those LIGO then detected later that year. Our estimates for the rate of merger events within primordial black hole clusters fit perfectly within the limits set by LIGO. If LIGO and other similar facilities detect many more mergers within the next few years, it may be possible to deter-

Illustration by Jen Christiansen and George Retseck (density-fluctuation panels)

© 2018 Scientific American

Is Dark Matter Made of Primordial Black Holes?

The quest for observational evidence continues

Seven observations point toward the hypothesis that primordial black holes constitute dark matter in the universe:

1. Mass, spin and merging rates of black holes. Gravitational-wave detectors are finding black hole mergers at a rate suggesting an abundance of black holes similar to the dark matter. In June 2017 LIGO announced the detection of black hole merger GW170104, for which the axis of rotation of one of the black hole progenitors was very likely not aligned with the axis of the orbital plane. This would be a sign that the black hole binary probably formed by a capture process, as expected for primordial black holes rather than from a preexisting star binary. Ultimately if we found a black hole with a mass lower than the so-called Chandrasekhar limit (1.4 solar masses, below which stars cannot produce a black hole), it would undeniably have a primordial origin. Fortunately, LIGO may very soon reach the sensitivity to detect such a black hole if its companion is more massive (greater than 10 solar masses). Finally, on cosmological scales, abundant black hole binaries should induce a diffuse background of gravitational waves, which could be detected by the future space-based Laser Interferometer Space Antenna (LISA) and by ground-based pulsar timing arrays.

2. Microlensing of stars in M31 and distant quasars. A broad mass distribution of primordial black holes, centered on a few solar masses, would perfectly explain the LIGO/VIRGO gravitational-wave detections but also the microlensing of stars in the Andromeda galaxy (M31) and of distant quasars if they constitute dark matter. Such a model passes all the other cosmological and astrophysical constraints.

3. Ultrafaint dwarf galaxies. In 2015 astronomers using data from the Dark Energy Survey discovered dozens of ultrafaint dwarf galaxies in the galactic halo, suggesting that hundreds of such dark matter–dominated dwarf galaxies could orbit around the Milky Way. Strangely, none of them had a radius smaller than 50 light-years. Regular close encounters of primordial black holes with stars would make such dwarf galaxies unstable. On the other hand, a rapid accretion episode in the densest halos provides a new mechanism to explicate why some dwarf galaxies are strongly dominated by dark matter and others are not. Future space-based facilities such as the European Space Agency's Euclid mission and NASA's Wide-field Infrared Survey Telescope (WFIRST) should detect many more ultrafaint dwarf galaxies and so will test this hypothesis.

4. Galactic dark matter profiles. New observations suggest the dark matter distribution is more homogeneous near the center of galaxies of all sizes, which would hardly be explained by any known astrophysical processes such as supernovae feedback. Primordial black holes of several solar masses would naturally explain the existence of core dark matter profiles, thanks to gravitational scattering that homogenizes their distribution in the densest regions.

5. Correlations between x-ray and infrared backgrounds. Alexander Kashlinsky of NASA's Goddard Space Flight Center suggested that correlations between x-ray and infrared cosmic background fluctuations that cannot be explained by any known population of galaxies would naturally arise if dark matter is made of massive primordial black holes. They would emit x-rays when they accrete matter within black hole clusters where a first generation of stars could also ignite and shine in infrared.

6. Billions of supermassive black holes. Recent observations of Chandra suggest that billions of supermassive black holes already existed one billion years after the big bang, shortly after the formation of the first stars. The required black hole seeds could have a primordial origin.

7. Unexpectedly cool universe at the time of star formation. In February 2018 a modest radio telescope built by a small team of astronomers at the Massachusetts Institute of Technology and at Arizona State University, called EDGES, might have detected the light from the epoch of formation of the first star generation, 200 million years after the big bang. Yet the amplitude of the signal was totally unexpected and suggests that the universe was half as warm—or colder—as in any realistic scenario. Primordial black hole dark matter might be the most natural explanation for this unexpected chilliness: the neutral hydrogen gas could have been extradiluted and thus extracooled because a fraction of it accreted onto black holes and simply disappeared from our universe.

mine the range of masses and spins for all the progenitor black holes. Such a statistical analysis of black hole mergers would provide crucial information for testing their potentially primordial origins.

A key aspect of this scenario is that it evades the constraints on MACHOs previously set by gravitational microlensing experiments—constraints that ruled out black holes of up to about 10 solar masses as the main constituent of dark matter. If primordial black holes exist and possess a wide range of masses, only a small fraction would be visible to these microlensing experiments, with the bulk remaining invisible. And if primordial black holes are grouped in clusters, this arrangement suggests a probability of less than one in 1,000 that a cluster would happen to be along the line of sight of the stars in the nearby satellite galaxies monitored for microlensing events.

To avoid this effect, one could search for microlensing events elsewhere in the sky, looking for the magnified light from stars in the Milky Way's neighboring Andromeda galaxy or even from quasars in far distant galaxies. In this way, one could probe a much larger volume of galactic halos for signs of MACHOs—that is, for primordial black holes. Recent observations suggest that whereas MACHOs of up to 10 solar masses may not make up the entirety of an average galaxy's halo, MACHOs between one tenth and a few solar masses could easily account for about 20 percent of the mass in a typical galactic halo. This value is consistent with our broad-mass primordial black hole scenario.

Simply put, we cannot yet rule out the possibility that dark matter is mostly made up of primordial black holes. Indeed, this proposed scenario could decipher several other cosmic mysteries related to dark matter and galaxy formation.

© 2018 Scientific American

MANY PROBLEMS, ONE SOLUTION

CLUSTERS OF PRIMORDIAL BLACK HOLES could clear up the so-called missing satellite problem—the apparent lack of "dwarf" satellite galaxies that should form around massive galaxies such as our Milky Way. Current simulations modeling the cosmic distribution of dark matter accurately replicate the universe's observed large-scale structure, in which halos of dark matter pull galaxy clusters into giant filaments and sheets surrounding great voids of lower density. On smaller scales, however, these simulations predict the existence of numerous subhalos of dark matter orbiting around massive galaxies. Each of these subhalos should host a dwarf galaxy, and hundreds should surround the Milky Way. Yet astronomers have found far fewer dwarf galaxies than predicted.

Many potential explanations for the missing satellite problem exist, mainly the notion that simulations fail to fully account for the influence of ordinary matter (hydrogen and helium in stars) on the formation and behavior of the predicted dwarf galaxies. Our scenario suggests that if clustered primordial black holes made up most dark matter, they would dominate the subhalos surrounding the Milky Way, absorbing a fraction of ordinary matter and reducing the rate of star formation in the subhalos. Moreover, even if these subhalos vigorously formed stars, these stars could easily be ejected by close encounters with massive primordial black holes. Both effects would greatly reduce the brightness of the satellites, making them very hard to detect without wide-field cameras of exquisite sensitivity. Fortunately, such cameras now exist, and astronomers have already used them to discover dozens of ultrafaint dwarf galaxies surrounding the Milky Way. These objects appear to host up to hundreds of times more dark matter than luminous stars, and our model predicts that thousands more should orbit our galaxy.

Simulations also predict a population of galaxies intermediate in size between dwarf galaxies and massive galaxies. Such objects are said to be too big to fail because they would be sufficiently large to readily form stars and be easily seen. Still, they have not turned up in astronomers' searches of the Milky Way's vicinity. This too-big-to-fail problem has a solution similar to that of the missing satellite problem: massive primordial black holes in the cores of intermediate-sized galaxies could eject stars and star-forming gas from these objects, rendering them effectively invisible to most surveys.

Primordial black holes could also resolve the origin of supermassive black holes (SMBHs). These monsters weigh from millions to billions of solar masses and are observed at the centers of quasars and massive galaxies very early in the universe's history. Yet if these SMBHs formed and grew from the gravitational collapse of the universe's first stars, they should not have acquired such gigantic masses in such a relatively short time—less than a billion years after the big bang.

In our scenario, although most primordial black holes have just tens of solar masses, a very small fraction will be far heavier, ranging from hundreds to tens of thousands of solar masses. Born less than a second after the big bang, these monstrous objects would then act as giant seeds for the formation of the first galaxies and quasars, which would rapidly develop SMBHs at their centers. Such seeds could also account for the existence of intermediate-mass black holes possessing 1,000 to a million solar masses observed orbiting SMBHs and at the centers of globular clusters of stars. In short, primordial black holes may be the missing link between conventional stellar-mass black holes and SMBHs. The observational case for this scenario is building rapidly: recent detections of unexpectedly abundant x-ray sources in the early universe are most easily explained by large numbers of primordial black holes producing x-rays as they gorge on gas less than one billion years after the big bang.

SEEING IN THE DARK

EVEN THOUGH massive primordial black holes could solve the mystery of dark matter, as well as many other long-standing problems of cosmology, the game is not yet over. Other models and explanations are still possible, and future observations should allow us to distinguish among the alternatives. Indeed, there already exist seven kinds of observations to test the primordial black hole scenario [*see box on opposite page*]. They include the detection of ultrafaint dwarf galaxies, the influence of massive primordial black holes on the positions of stars in the Milky Way, the mapping of neutral hydrogen during the first epoch of star formation and future observations of distortions in the cosmic microwave background.

Beyond these experiments, we also now possess a completely new tool to unravel the mysteries of the universe in the form of Advanced LIGO and other gravitational-wave detectors. If indeed LIGO has detected merging members of a hidden population of massive primordial black holes, we should expect many more to be detected in coming years. Scientists from the Advanced LIGO and VIRGO collaborations have presented so far five confirmed detections of gravitational waves emitted during the merging of black hole binaries, and several new detections will be soon announced to the community. These detections suggest that binary black holes are much more frequent than expected and that they are broadly distributed in mass, in agreement with our scenario of massive primordial black holes.

Taken together these new experiments and observations could confirm the existence of primordial black holes and their possible linkage to the universe's missing matter. Soon we may no longer be in the dark about dark matter. SA

Juan García-Bellido is a theoretical physicist and professor at the Institute for Theoretical Physics in Madrid. His research interests are the early universe, dark energy, black holes and quantum gravity. García-Bellido is an active member of the Dark Energy Survey and the European Space Agency missions Euclid and Laser Interferometer Space Antenna (LISA).

Sébastien Clesse is a Belgian cosmologist and an F.R.S.-FNRS postdoctoral researcher at the University of Louvain and University of Namur in Belgium. His work covers cosmic inflation, modified gravity and primordial black holes. Clesse is an active member of the Euclid mission and of the Square Kilometer Array collaboration.

MORE TO EXPLORE

Density Perturbations and Black Hole Formation in Hybrid Inflation. Juan García-Bellido, Andrei Linde and David Wands in *Physical Review D,* Vol. 54, No. 10, pages 6040–6058; November 15, 1996.

LIGO Gravitational Wave Detection, Primordial Black Holes, and the Near-IR Cosmic Infrared Background Anisotropies. A. Kashlinsky in *Astrophysical Journal Letters,* Vol. 823, No. 2, Article No. L25; June 1, 2016.

The Clustering of Massive Primordial Black Holes as Dark Matter: Measuring Their Mass Distribution with Advanced LIGO. Sébastien Clesse and Juan García-Bellido in *Physics of the Dark Universe,* Vol. 15, pages 142–147; March 2017.

scientificamerican.com/magazine/sa

© 2018 Scientific American

The Exer*Paradox*

Studies of how the human engine burns calories help to explain why physical activity does little to control weight—and how our species acquired some of its most distinctive traits

By Herman Pontzer

Illustration by Bomboland

IN BRIEF

Conventional wisdom holds that physically active people burn more calories than less active people do.
But studies show that traditional hunter-gatherers, who lead physically hard lives, burn the same number of calories as people with access to modern conveniences.
The discovery that human energy expenditure is tightly constrained raises questions about how our large brain and other energetically demanding traits evolved.
Comparisons with energy expenditure in great apes suggest that the human metabolic engine has evolved to get more work done to support our costly features.

© 2018 Scientific American

rcise

REVOLUTIONS IN SCIENCE

HEALTH

© 2018 Scientific American

Still no giraffe.

Four of us had been walking half the day, tracking a wounded giraffe that Mwasad, a Hadza man in his late 30s, shot the evening before. He hit it in the base of the neck from about 25 yards with a steel-tipped, wood arrow smeared with powerful, homemade poison. Hadza are traditional hunter-gatherers who live off wild plants and animals in the dry savanna wilderness of northern Tanzania. They know the landscape and its residents better than you know your local Trader Joe's. Mwasad had let the giraffe run to give the poison time to work, hoping to find it dead in the morning. An animal that size would feed his family and his camp for a week—but only if he could locate it.

Mwasad led our party—Dave Raichlen from the University of Arizona, a 12-year-old Hadza boy named Neje and me—out of camp just after daybreak. Dave and I were of little use in this endeavor. Mwasad had invited us along as a friendly gesture and for the extra help to carry the butchered animal back to camp should our search effort succeed. As anthropologists who study human ecology and evolution, we jumped at the opportunity to tag along—Hadza men's tracking abilities are legendary. It certainly beat the prospect of a long day in camp spent fiddling with research equipment.

We walked hard for an hour through a pathless, rolling sea of golden, waist-high grass, dotted with brush and thorny acacia trees, directly to the bloody patch where the giraffe was struck. That bit of navigation in itself was quite a trick, like someone leading you to the middle of a 1,000-acre wheat field to show you where he had once dropped a toothpick and then nonchalantly reaching down to pick it up. Hour on hour, tracking the wounded animal under a relentless sun ensued as we followed ever more tenuous signs.

Still no giraffe. At least I had water. We sat in the shade of some bushes just after midday, taking a break while Mwasad pondered where the injured creature might head. I had a quart or so left—enough, I figured, to get through the heat of the afternoon. Mwasad, however, had not brought any water with him, as is typical of the Hadza. As we packed up to restart the search, I offered him a drink. Mwasad gave me a sideways look, smiled and proceeded to drink the entire bottle in one long pull. When he finished, he casually handed me the empty bottle.

It was karma. Dave and I, along with anthropologist Brian Wood, then at Yale University, had spent the past month living with the Hadza, conducting the first direct measurements of daily energy expenditure in a hunter-gatherer population. We enlisted a couple of dozen Hadza women and men, Mwasad among them, to drink small, incredibly expensive bottles of water enriched in two rare isotopes, deuterium and oxygen 18. Analyzing the concentration of those isotopes in urine samples from each participant would allow us to calculate their body's daily rate of carbon dioxide production and thus their daily energy expenditure. This approach, known as the doubly labeled water method, is the gold standard in public health for measuring the calories burned each day during normal daily life. It is straightforward, completely safe and accurate, but it requires that participants drink every last drop of the enriched water. We had taken pains to make clear that they must not spill, that they had to finish the dose completely. Mwasad seemed to have taken that message to heart.

Mwasad's sly joking aside, my colleagues and I have learned a lot about how the human body burns calories through our work with the Hadza. Together with findings from investigators who study other populations, our research has revealed some surprising insights into human metabolism. Our data indicate that, contrary to received wisdom, humans tend to burn the same number of calories regardless of how physically active they are. Yet we have evolved to burn considerably more calories than our primate cousins do. These results help to explain two puzzles that might seem disparate at first but are, in fact, related: first, why exercise generally fails to aid weight loss and, second, how some of humanity's unique traits arose.

THE CALORIE ECONOMY

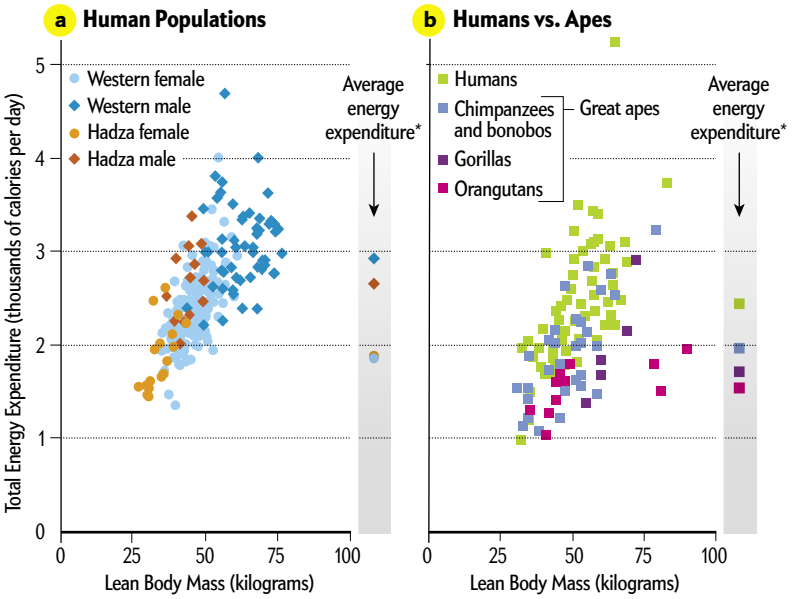
RESEARCHERS WHO ARE INTERESTED in human evolution and ecology often focus on energy expenditure because energy is central to everything in biology. One can learn a lot about any species by measuring its metabolism: life is essentially a game of turning energy into kids, and every trait is tuned by natural selection to maximize the evolutionary return on each calorie spent. Ideally, the study population lives in the same environments in which the species originally evolved, where the same ecological pressures that shaped its biology are still at work. That is difficult to achieve with human subjects because most people are divorced from the daily work of acquiring food from a wild landscape. For nearly all the past two million years, humans and our ancestors have been living and evolving as hunter-gatherers. Farming only got going about 10,000 years ago; industrialized cities and modern technology are only a few generations old. Populations such as the Hadza, one of the last hunter-gatherer populations left in the world, are key to understanding how our bodies evolved and functioned before cows, cars and computers.

© 2018 Scientific American

Gas Guzzlers

Experts have assumed that physically active people burn more calories than less active folks. But direct measurements of energy expenditure in modern-day hunter-gatherers in the developing world and comparatively sedentary populations in the U.S. and Europe reveal similar results **a**. If human metabolism is so tightly constrained, how did our big brain, long life span and other energetically costly features that distinguish us from our primate relatives evolve? Humans consume and expend hundreds of calories a day more than great apes **b**, suggesting that our metabolic engine changed to burn energy faster, thus powering our expensive traits.

**To account for differences in energy expenditure arising from body size, Western averages are calculated at Hadza body sizes, and great ape averages are calculated at average human body size.*



Life for the Hadza is physically demanding. Each morning the women leave the grass huts of camp in small groups, some carrying infants on their back in a wrap, foraging for wild berries or other edibles. Wild tubers are a staple of the Hadza diet, and women can spend hours digging them out of the rocky ground with sticks. Men cover miles each day hunting with bows and arrows they make themselves. When game is scarce, they use simple hatchets to chop into tree limbs, often 40 feet up in the canopy, to harvest wild honey. Even the children contribute, hauling buckets of water back from the nearest watering hole, sometimes a mile or more from camp. In the late afternoon, folks wend their way back to camp, sitting on the ground and talking around small cooking fires, sharing the day's returns and tending to the kids. Days roll along like this through dry and wet seasons, ad millennium.

But forget any romantic notions of some lost Eden. Hunting and gathering is cerebral and risky, a high-stakes game in which the currency is calories and going bust means death. Men such as Mwasad spend hundreds of calories a day hunting and tracking, a gamble that they hope will pay off in game. Savvy is just as critical as stamina. Whereas other predators can rely on their speed and strength to obtain prey, humans have to outthink their quarry, considering their behavioral tendencies and scouring the landscape for signs of game. Still, Hadza men land big game like giraffes only about once a month. They would starve if Hadza women were not executing an equally sophisticated, complementary strategy, using their encyclopedic knowledge of local plant life to bring home a reliable bounty every day. This complex, cooperative foraging is what made humans so incredibly successful and is the core of what makes us unique.

Researchers in public health and human evolution have long assumed that our hunter-gatherer ancestors burned more calories than people in cities and towns do today. Given how physically hard folks such as the Hadza work, it seems impossible to imagine otherwise. Many in public health go so far as to argue that this reduction in daily energy expenditure is behind the global obesity pandemic in the developed world, with all those unburned calories slowly accumulating as fat. One of our motivations for measuring Hadza metabolism was to determine the size of this energy shortfall and see just how deficient we Westerners were in our daily expenditure. Back home in the U.S. after a hot and dusty field season, I lovingly packed the vials of Hadza urine on dry ice and sent them away to the Baylor College of Medicine, home of one of the best doubly labeled water laboratories in the country, imagining the whopping calorie totals they would reveal.

But a funny thing happened on the way to the isotope ratio mass spectrometer. When the analyses came back from Baylor, the Hadza looked like everyone else. Hadza men ate and burned about 2,600 calories a day, Hadza women about 1,900 calories a day—the same as adults in the U.S. or Europe. We looked at the data every way imaginable, accounting for effects of body size, fat percentage, age and sex. No difference. How was it possible? What were we missing? What else were we getting wrong about human biology and evolution?

LIES MY FITBIT TOLD ME

IT SEEMS SO OBVIOUS and inescapable that physically active people burn more calories that we accept this paradigm without much critical reflection or experimental evidence. But since the 1980s and 1990s, with the advent of the doubly labeled water method, the empirical data have often challenged the conventional wisdom in public health and nutrition. The Hadza result, strange as it seemed, was not some thunderbolt from the blue but more like the first cold drop of water down your neck from a rain that had been building, ignored, for years.

The earliest doubly labeled water studies among traditional farmers in Guatemala, the Gambia and Bolivia showed their energy expenditures were broadly similar to those of city dwellers. In a study published in 2008, Amy Luke, a researcher in public health at Loyola University Chicago, took this work a step further, comparing energy expenditure and physical activity in

SOURCES: "HUNTER-GATHERER ENERGETICS AND HUMAN OBESITY," BY HERMAN PONTZER ET AL., IN *PLOS ONE*, VOL. 7, NO. 7, ARTICLE NO. E40503; JULY 25, 2012 *(a)*; "METABOLIC ACCELERATION AND THE EVOLUTION OF HUMAN BRAIN SIZE AND LIFE HISTORY," BY HERMAN PONTZER ET AL., IN *NATURE*, VOL. 533; MAY 19, 2016 *(b)*

Graphic by Jen Christiansen

© 2018 Scientific American

rural Nigerian women with those measures in African-American women in Chicago. Like the Hadza study, hers found no differences in daily energy expenditure between populations, despite large differences in activity levels. Following up on that work, Lara Dugas, also at Loyola, along with Luke and others, analyzed data from 98 studies around the globe and showed that populations coddled by the modern conveniences of the developed world have similar energy expenditures to those in less developed countries, with more physically demanding lives.

Humans are not the only species with a fixed rate of energy expenditure. On the heels of the Hadza study, I piloted a large collaborative effort to measure daily energy expenditure among primates, the group of mammals that includes monkeys, apes, lemurs and us. We found that captive primates living in labs and zoos expend the same number of calories each day as those in the wild, despite obvious differences in physical activity. In 2013 Australian researchers found similar energy expenditures in sheep and kangaroos kept penned or allowed to roam free. And in 2015 a Chinese team reported similar energy expenditures for giant pandas in zoos and in the wild.

For a more granular look, comparing individuals instead of population averages, I joined Luke and her team, including Dugas, to examine activity and energy expenditure in a large, multiyear analysis known as the Modeling the Epidemiological Transition Study (METS). More than 300 participants wore accelerometers similar to a Fitbit or other fitness tracker 24 hours a day for an entire week while their daily energy expenditure was measured with doubly labeled water. We found that daily physical activity, tracked by the accelerometers, was only weakly related to metabolism. On average, couch potatoes tended to spend about 200 fewer calories each day than people who were moderately active: the kind of folks who get some exercise during the week and make a point to take the stairs. But more important, energy expenditure plateaued at higher activity levels: people with the most intensely active daily lives burned the same number of calories each day as those with moderately active lives. The same phenomenon keeping Hadza energy expenditure in line with that of other populations was evident among individuals in the study.

How does the body adjust to higher activity levels to keep daily energy expenditure in check? How can the Hadza spend hundreds of calories a day on activity yet burn the same total number of calories a day as comparatively sedentary people in the U.S. and Europe? We are still not sure, but the cost of activity per se is not changing: we know, for example, that Hadza adults burn the same number of calories to walk a mile as Westerners do. It could be that people with high activity levels change their behavior in subtle ways that save energy, like sitting rather than standing or sleeping more soundly. But our analysis of the METS data suggests that although these behavioral changes might contribute, they are not sufficient to account for the constancy seen in daily energy expenditure.

Another intriguing possibility is that the body makes room for the cost of additional activity by reducing the calories spent on the many unseen tasks that take up most of our daily energy budget: the housekeeping work that our cells and organs do to keep us alive. Saving energy on these processes could make room in our daily energy budget, allowing us to spend more on physical activity without increasing total calories spent per day.

HADZA HUNTER-GATHERERS in Tanzania spend hundreds of calories a day on activity yet burn the same total number of calories as city dwellers in the U.S.

HARRY HOOK Getty Images

For example, exercise often reduces inflammatory activity that the immune system mounts, as well as levels of reproductive hormones such as estrogen. In lab animals, increased daily exercise has no effect on daily energy expenditure but instead results in fewer ovulatory cycles and slower tissue repair. And extremes may lead some animals to eat their own nursing infants. Humans and other creatures seem to have several evolved strategies for keeping daily energy expenditure constrained.

All of this evidence points toward obesity being a disease of gluttony rather than sloth. People gain weight when the calories they eat exceed the calories they expend. If daily energy expenditure has not changed over the course of human history, the primary culprit in the modern obesity pandemic must be the calories consumed. This should not be news. The old adage in public health is that "you can't outrun a bad diet," and experts know from personal experience and lots of data that just hitting the gym to lose weight is frustratingly ineffective. But the new science helps to explain why exercise is such a poor tool for weight loss. It is not that we are not trying hard enough. Our bodies have been plotting against us from the start.

You still have to exercise. This article is not a note from your mom excusing you from gym class. Exercise has tons of well-documented benefits, from increased heart and immune system health to improved brain function and healthier aging. In fact, I suspect that metabolic adaptation to activity is one of the reasons exercise keeps us healthy, diverting energy away from activities, such as inflammation, that have negative consequences if they go on too long. For example, chronic inflammation has been linked to cardiovascular disease and autoimmune disorders.

The foods we eat certainly affect our health, and exercise paired with dietary changes can help keep off unwanted pounds once a healthy weight has been reached, but evidence indicates that it is best to think of diet and exercise as different tools with different strengths. Exercise to stay healthy and vital; focus on diet to look after your weight.

© 2018 Scientific American

ENERGY BUDGETS AND EVOLUTION

EVEN AS THE RECENT SCIENCE in metabolic adaptation helps to clarify the relation between exercise and obesity, a constrained, adaptive metabolism leaves researchers with larger, existential questions. If daily energy expenditure is virtually immobile, how could humans evolve to be so radically different from our ape relatives? Nothing in life is free. Resources are limited, and investing more in one trait inevitably means investing less in another. It is no coincidence that rabbits reproduce prodigiously but die young; all that energy plowed into offspring means less for bodily maintenance and longevity. *Tyrannosaurus rex* can thank its big head of nasty teeth and powerful hind limbs for its puny arms and hands. Even dinosaurs couldn't have it all.

Humans flout this bedrock evolutionary principle of austerity. Our brains are so large that, as you sit reading this article, the oxygen from every fourth breath you take is needed just to feed your brain. Yet humans have bigger babies, reproduce more often, live longer and are more physically active than any of our ape relatives. Hadza camps are full of cheerfully chaotic children and hale, hearty men and women in their 60s and 70s. Our energetic extravagance presents an evolutionary puzzle. Humans are so genetically and biologically similar to other apes that researchers have long assumed that our metabolisms are similar, too. But if energy expenditures are as constrained as our Hadza study and others suggest, how could an inflexible, apelike metabolism process all the calories needed to support our costly human traits?

In the wake of our broad, comparative primate energetics study, my colleagues and I began to wonder whether humans' adaptive suite of energetically costly traits was fueled by a wholesale evolutionary change in metabolic physiology. We had found in that study that primates burn only half as many calories a day as other mammals do. The reduced metabolic rates of primates correspond with their slow rates of growth and reproduction. Perhaps, conversely, the faster reproduction and other expensive traits of humans were linked to the evolution of an increased metabolic rate. All that was needed to test this idea was getting a bunch of frenetic chimpanzees, wily bonobos, phlegmatic orangutans and skittish silverback gorillas to carefully drink doses of doubly labeled water without spilling and to provide a few urine samples. In a scientific tour de force, my colleagues Steve Ross and Mary Brown, both at Lincoln Park Zoo in Chicago, worked with caretakers and veterinarians from more than a dozen zoos across the U.S. to pull that off. It took a couple of years, but they accumulated enough data on great ape energy expenditure to provide a solid comparison with humans.

Sure enough, humans burn more calories each day than any of our great ape relatives. Even after accounting for effects of body size, activity level and other factors, humans consume and expend about 400 more calories a day than chimpanzees and bonobos do; differences with gorillas and orangutans are larger still. Those extra calories represent the extra work our bodies do to support larger brains, produce more babies and maintain our bodies so we live longer. It is not simply that we eat more than other apes (although we do that, too); as we know all too well, piling extra calories into a body that is not equipped to use them only results in obesity. Our bodies, right down to the cellular level, have evolved to burn energy faster and get more done than our ape relatives. Human evolution was not entirely without trade-offs: our digestive tract is smaller and less costly than other apes, which need a large, energetically expensive gut to digest their fibrous, plant-based diets. But the critical changes that make us human were fueled by an evolutionary shift in our metabolic engine.

SHARED FORTUNES

AT SOME POINT in the late afternoon, our path bent toward camp, Mwasad looking ahead instead of searching the ground. We were heading home sans giraffe. Here was the fundamental danger in the high-energy human strategy: coming home empty-handed was both more likely and more consequential. Many of the energy-rich foods we need to fuel our faster metabolisms are inherently difficult to obtain in the wild, increasing the energy cost of finding food and heightening the risk of starvation for the men and women out foraging and their kids back at camp.

Happily for Mwasad, humans have evolved a few tricks to keep starvation at bay. We are the only species that has learned to cook, which increases the caloric value of many foods and makes them more efficient to digest. Our mastery of fire converts otherwise inedible root vegetables—from Trader Joe's yams to wild Hadza tubers—into veritable starch bombs. We have also evolved to be fat. We know this all too well from obesity rates in the West, but even Hadza adults, lean by any human standard, carry twice as much fat as chimpanzees idling away in zoos. Problematic though it may be in our modern era, our propensity to store fat most likely coevolved with our faster metabolism as a critical energy buffer to survive lean times.

As the sun sat heavy and orange just above the trees, we melted back into camp, Dave and I toward our tents, Mwasad and Neje to their families' huts, each one of us glad to be home. Despite the lost giraffe, no one went hungry that evening. Instead, with little fanfare or conscious effort, the camp deployed our species' most ingenious and powerful weapon against starvation: sharing. Sharing food is so fundamental to the human experience, the common thread of every barbecue, birthday, bar mitzvah, that we take it for granted, but it is a unique and essential part of our evolutionary inheritance. Other apes do not share.

Beyond our nutritional requirements and fixation with fat, perhaps the most profound impact of our increased energy expenditure is this human imperative to work together. Evolving a faster metabolism bound our fortunes to one another, requiring that we cooperate or die. As I sat with Dave and Brian, recounting the day's adventures over tinned sardines and potato chips, I realized I would not have had it any other way. No giraffe, no problem. SA

Herman Pontzer is an anthropologist at Hunter College. He studies energy expenditure in humans and great apes to test hypotheses about the evolution of human physiology and anatomy.

MORE TO EXPLORE

Hunter-Gatherer Energetics and Human Obesity. Herman Pontzer et al. in *PLOS ONE,* Vol. 7, No. 7, Article No. e40503; July 25, 2012.

Constrained Total Energy Expenditure and Metabolic Adaptation to Physical Activity in Adult Humans. Herman Pontzer et al. in *Current Biology,* Vol. 26, No. 3, pages 410–417; February 8, 2016.

Metabolic Acceleration and the Evolution of Human Brain Size and Life History. Herman Pontzer et al. in *Nature,* Vol. 533, pages 390–392; May 19, 2016.

scientificamerican.com/magazine/sa

© 2018 Scientific American

THE NEW MIGRAINE MEDS

The cause of migraine headaches has eluded scientists for centuries. Now a theory blaming one nerve has led to drugs that prevent attacks

By David Noonan

© 2018 Scientific American

Illustration by Julia Yellow

© 2018 Scientific American

THE 63-YEAR-OLD CHIEF EXECUTIVE COULDN'T DO HIS JOB. HE HAD BEEN CRIPPLED by migraine headaches throughout his adult life and was in the middle of a new string of attacks. "I have but a little moment in the morning in which I can either read, write or think," he wrote to a friend. After that, he had to shut himself up in a dark room until night. So President Thomas Jefferson, in the early spring of 1807, during his second term in office, was incapacitated every afternoon by the most common neurological disability in the world. The co-author of the Declaration of Independence never vanquished what he called his "periodical head-ach," although his attacks appear to have lessened after 1808.

Two centuries later 36 million American migraine sufferers grapple with the pain the president felt. Like Jefferson, who often treated himself with a concoction brewed from tree bark that contained quinine, they try different therapies, ranging from heart drugs to yoga to herbal remedies. Their quest goes on because modern medicine, repeatedly baffled in attempts to find the cause of migraine, has struggled to provide reliable relief.

Now a new chapter in the long and often curious history of migraine is being written. Neurologists believe they have identified a hypersensitive nerve system that triggers the pain and are in the final stages of testing medicines that soothe its overly active cells. These are the first ever treatments specifically designed to prevent the crippling headaches before they start. If they deliver on the promise they have shown in studies conducted so far, including two recent phase III clinical trials, millions of headaches may never happen.

"It completely changes the paradigm of how we treat migraine," says David Dodick, a neurologist at the Mayo Clinic's campus in Arizona and a past president of the International Headache Society. Whereas there are migraine-specific drugs that do a good job stopping attacks after they start, the holy grail for both patients and doctors has been prevention.

Migraine attacks, which affect almost 730 million people worldwide, typically last from four to 72 hours. Most sufferers have sporadic migraines and are laid low during 14 or fewer days a month. Those with a chronic form—almost 8 percent of the migraine population—suffer 15 or more monthly "headache days." Attacks are often preceded by fatigue, mood changes, nausea and other symptoms. About 30 percent of migraine patients experience visual disturbances, called auras, before headaches hit. The total economic burden of migraine in the U.S., including direct medical costs and indirect costs such as lost workdays, is estimated at $17 billion annually.

In the 5,000 years since migraine symptoms were first described in Babylonian documents, treatments have reflected both our evolving grasp and our almost comical ignorance of the condition. Bloodletting, trepanation and cauterization of the shaved scalp with a red-hot iron bar were common treatments during the Greco-Roman period. The nadir of misguided remedies was probably reached in the 10th century A.D., when the otherwise discerning ophthalmologist Ali ibn Isa recommended binding a dead mole to the head. In the 19th century medical electricity had become all the rage, and migraine patients were routinely jolted with a variety of inventions, including the hydroelectric bath, which was basically an electrified tub of water.

By the early 20th century clinicians turned their attention to the role of the blood vessels, inspired in part by observations of strong pulsing of the temporal arteries in migraine patients, as well as patients' descriptions of throbbing pain and the relief they got from compression of the carotid arteries. For decades to come, migraine pain would be blamed primarily

IN BRIEF

Migraine headaches affect nearly 730 million people worldwide and result in a total economic burden of approximately $17 billion every year, and yet the cause of the condition has eluded scientists for hundreds of years.

In the 1980s researchers began to target the trigeminal nerve system, the brain's primary pain pathway. Animal studies had shown that overactive cells on that pathway would release pain signals and cause migraine symptoms.

Blocking one chemical found in high concentrations in the blood of people experiencing migraines—CGRP—is showing great promise as a potential treatment; several clinical trials of the tactic are showing better efficacy than any other available remedy.

© 2018 Scientific American

on the dilation of blood vessels (vasodilation) in the brain.

That idea was reinforced in the late 1930s with the publication of a paper on the use of ergotamine tartrate, an alkaloid that was known to constrict blood vessels. Despite an array of side effects, among them vomiting and drug dependence, it did stop attacks in a number of patients.

But if vasodilation was part of the puzzle, it was not the only thing going on in the brains of migraine sufferers, as the next wave of treatments suggested. In the 1970s cardiac patients who also had migraines started telling their doctors that the beta blockers they were taking to slow rapid heartbeats also reduced the frequency of their attacks. And migraine sufferers taking medicines for epilepsy and depression, and others receiving cosmetic Botox injections, reported relief. So headache specialists began prescribing these "borrowed" drugs for migraines. Five of the medications eventually were approved by the FDA for the condition. Unfortunately, it is still not known exactly how the adopted drugs (which are effective in only about 45 percent of cases and come with an array of side effects) help migraines. Dodick says they may act at various levels of the brain and brain stem to reduce excitability of the cortex and pain-transmission pathways.

BLOODLETTING, TREPANATION AND CAUTERIZATION OF THE SHAVED SCALP WITH A RED-HOT IRON BAR WERE COMMON TREATMENTS DURING THE GRECO-ROMAN PERIOD.

The first migraine-specific drugs, the triptans, were introduced in the 1990s. Richard Lipton, director of the Montefiore Headache Center in New York City, says triptans were developed in response to the older idea that the dilation of blood vessels is the primary cause of migraine; triptans were supposed to inhibit it. Ironically, subsequent drug studies show that they actually disrupt the transmission of pain signals in the brain and that constricting blood vessels is not essential. "But they work anyway," Lipton says. A survey of 133 detailed triptan studies found that they relieved headache within two hours in 42 to 76 percent of patients. People take them to stop attacks after they start, and they have become a reliable frontline treatment for millions.

What triptans cannot do—and what Peter Goadsby, a neurologist at the Headache Center at the University of California, San Francisco, has dreamed about doing for more than 30 years—is prevent migraine attacks from happening in the first place. In the 1980s, in pursuit of this goal, Goadsby focused on the trigeminal nerve system, long known to be the brain's primary pain pathway. It was there, he suspected, that migraine did its dirty work. Studies in animals indicated that in branches of the nerve that exit from the back of the brain and wrap around various parts of the face and head, overactive cells would respond to typically benign lights, sounds and smells by releasing chemicals that transmit pain signals and cause migraine. The heightened sensitivity of these cells may be inherited; 80 percent of migraine sufferers have a family history of the disorder.

Goadsby co-authored his first paper on the subject in 1988, and other researchers, including Dodick, joined the effort. Their goal was to find a way to block the pain signals. One of the chemicals found in high levels in the blood of people experiencing migraine is calcitonin gene–related peptide (CGRP), a neurotransmitter that is released from one nerve cell and activates the next one in a nerve tract during an attack. Zeroing in on CGRP and interfering with it was hard. It was difficult to find a molecule that worked on that neurotransmitter and left other essential chemicals alone.

As biotech engineers' ability to control and design proteins improved, several pharmaceutical companies developed migraine-fighting monoclonal antibodies. These designer proteins bind tightly to CGRP molecules or their receptors on trigeminal nerve cells, preventing cell activation. The new drugs are "like precision-guided missiles," Dodick says. "They go straight to their targets."

It is that specificity, and the fact that scientists actually know how the drugs work, that has Dodick, Goadsby and others excited. In two placebo-controlled trials with a total of 380 people who had severe migraines up to 14 days per month, a single dose of a CGRP drug decreased headache days by more than 60 percent (63 percent in one study and 66 percent in the other). Phase III clinical trials of two drugs, described in November 2017 in the *New England Journal of Medicine,* confirmed the efficacy of the approach. The drugs, erenumab and fremanezumab, were tested on a total of 1,395 patients and significantly reduced migraine frequency. So far the CGRP drugs, which are given as injections, work better at prevention than any of the borrowed heart or epilepsy drugs and have far fewer side effects.

Migraine specialists are also exploring other treatments, including forehead and eyelid surgery to decompress branches of the trigeminal nerve, as well as transcranial magnetic stimulation (TMS), a noninvasive way of altering nerve cell activity.

Lipton says he has had some good results with TMS. He has also referred patients for surgical interventions but says the experience "has been disappointing," and he is not recommending it. For his part, Goadsby views surgeries and high-tech efforts as a kind of desperation: "They strike me as a cry for help. If we better understood migraine, we'd know better what to do."

Even though the cause now appears rooted in the trigeminal nerve system, the origin of its overactive cells is still a mystery, Goadsby says. "What's the nature of what you inherit when you inherit migraine?" he asks. "Why you, and why not me?" If researchers untangle the genetics of migraine, Jefferson's "periodical head-ach" may loosen its painful modern grip. SA

David Noonan is a freelance writer specializing in science and medicine.

MORE TO EXPLORE

A Controlled Trial of Erenumab for Episodic Migraine. P. J. Goadsby et al. in *New England Journal of Medicine,* Vol. 366, pages 2123–2132; November 30, 2017.

Fremanezumab for the Preventive Treatment of Chronic Migraine. S. D. Silberstein et al. in *New England Journal of Medicine,* Vol. 377, pages 2113–2122; November 30, 2017.

scientificamerican.com/magazine/sa

© 2018 Scientific American

© 2018 Scientific American

EBOLA'S SECOND COMING

Brain deficits and more torment many virus survivors in Liberia. The top suspects are hidden viral remnants and immune system overreactions

By Seema Yasmin

Illustration by Edel Rodriguez

© 2018 Scientific American

JOSEPHINE KARWAH STEPPED OUT OF THE EBOLA TREATMENT UNIT AND cradled her pregnant belly. She had hobbled into the white tent in Monrovia, Liberia, two weeks earlier, during August of 2014, her knees burning with pain and threatening to buckle every fourth step. Josephine's mother had died in this unit. Her body had been carried away in a white body bag that nurses had prepared with her name written neatly on the side. Her father, too, had died from Ebola, as did her aunt and uncle. But Josephine, though she got sick from the virus, lived. She and her unborn child were survivors, unlike 40 percent of the patients in the 2014–2016 African Ebola epidemic. Josephine decided she would name the baby Miracle.

Then the nightmares began. Back at home in her village, Smell No Taste, an hour's drive east of the Liberian capital, Josephine dreamed of the family members she had lost to Ebola and the horrors of the treatment unit. Throbbing headaches interrupted her dreams, and her hips and knees ached as she tried to fall back asleep. During the day she helped her older sister make soap to sell at the market. But her right eye burned, and her left eye made the world appear cloudy, as if drops of dew had settled on a camera lens. At the money changer's booth, she walked away with the wrong change, unable to recall how many Liberian dollars were in her purse when she left the house.

Josephine is one of 1,500 Ebola survivors in Liberia. Like Josephine, many today suffer memory loss, joint pains, muscle aches and eye problems. These are not isolated anecdotes and vague reports. In February 2016, reporting findings from the largest-ever study of Ebola survivors at a conference in Boston, Mosoka Fallah, an epidemiologist from Liberia, said more than half of the patients who lived through an acute attack later reported muscle and joint problems. Two thirds had neurological difficulties, and 60 percent reported eye problems approximately one year after Ebola infection. Although the World Health Organization declared the public emergency over in March 2016, now people are living with what doctors call post-Ebola syndrome.

Post-Ebola syndrome has been spotted before. After small virus outbreaks in East and Central Africa in the past 20 years, survivors suffered joint pains, muscle aches and eye problems serious enough to prevent many from working.

But these were limited episodes of the disease and small groups of survivors. The 2014–2016 West African Ebola epidemic has left 17,000 survivors at risk of post-Ebola syndrome. Like Josephine, they stepped out of treatment units into an uncertain future. There is one thing that experts and patients do know: Ebola is not over.

EBOLA'S GHOST

FALLAH'S OFFICE sits at one end of a long corridor in the John F. Kennedy Medical Center in Monrovia. A Harvard University–trained epidemiologist who grew up in one of Liberia's largest slums, he was part of the team testing treatments and vaccines as part of the initial Ebola response. His survivor research grew out of that work.

The National Institutes of Health in the U.S. and the Liberian Ministry of Health and Social Welfare had formed a coalition in 2014 called the Partnership for Research on Ebola Vaccines in Liberia (PREVAIL). By the time the initial vaccine safety tests were completed, however, Liberia's epidemic was slowing down. The number of people becoming infected with Ebola was far fewer than expected, so the first study, PREVAIL I, was scaled back to test only for vaccine safety and immune response and not the vaccine's ability to prevent Ebola. Instead PREVAIL scientists shifted resources to Ebola's aftereffects. Reports were coming in from across West Africa of patients who survived the disease but suffered physical and psychological problems. Fallah was appointed principal investigator for the study in Liberia and switched his focus from the Ebola response to Ebola survivors.

On a Wednesday afternoon, two days before Christmas, Fal-

IN BRIEF

With West Africa's Ebola epidemic declared over in 2016, about 17,000 people are at risk for symptoms called post-Ebola syndrome.

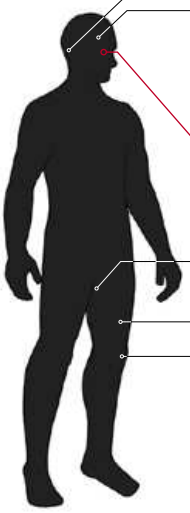
In a Liberian study, 60 percent of survivors reported eye problems, 53 percent had muscle aches and joint pain, and 68 percent had neurological problems.

Supposedly virus-free people are frequently shunned by others and must wonder if the disease will afflict them again.

© 2018 Scientific American

Suspected Hideouts

More than half of the people who live through Ebola have serious symptoms after the disease is supposedly gone, according to a study of more than 1,000 of Liberia's confirmed 1,500 virus survivors. Ailments include neurological problems and muscle pain. Physicians suspect the virus may still be hiding in parts of the body not rigorously patrolled by immune system cells, such as the eye. Or perhaps the initial immune reaction to the virus causes inflammation that damages organs. Here are commonly afflicted body parts and symptoms. The eye is a particular focus because about 60 percent of patients report troubles there.



Ears: Full or partial loss of hearing or a ringing sensation called tinnitus

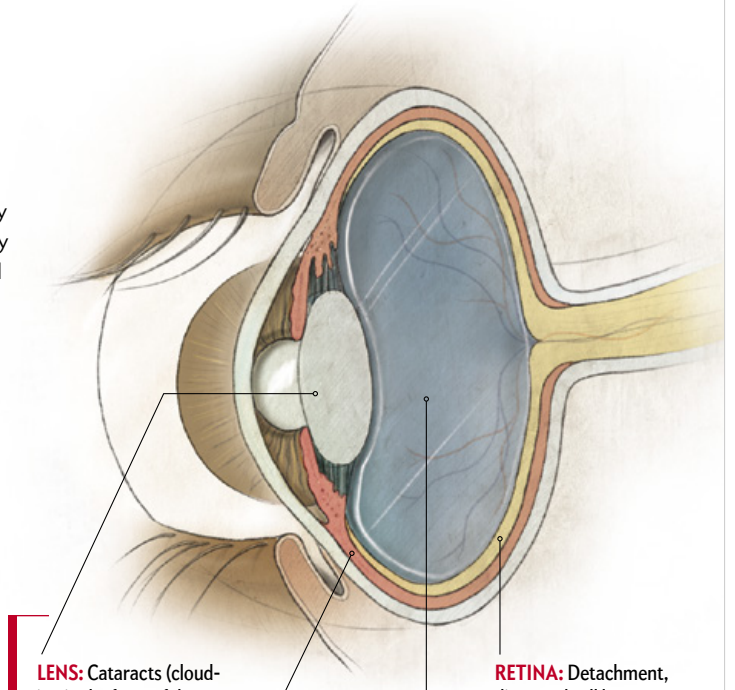
Brain: Headaches, movement-control issues resembling Parkinson's disease, memory problems, cognitive disorders, fatigue, anxiety, depression and sleep disorders
Suspected virus hiding place

Eyes:
Suspected virus hiding place

Testes:
Suspected virus hiding place

Muscles: Pain

Joints and joint cartilage: Pain
Suspected virus hiding place



LENS: Cataracts (clouding in the front of the lens instead of more typical obstructions in the middle or back), leading to blurred or lost vision, difficulty seeing in dim light, and halos around light sources

VITREOUS HUMOR: Protein clumping and particulate matter floating around, causing blurred vision

UVEA: Swelling of the middle layer, producing eye redness, pain, light sensitivity, blurred or decreased vision, and floating spots

RETINA: Detachment, disrupted cell layers, changes in pigment patterns and an inflamed area in the middle of the retina, possibly resulting in blindness or partial vision loss, inability to see in dim light, sensitivity to bright light, and seeing flashes or spots

lah flicked through a patient file at the Kennedy Medical Center. He had overseen the refurbishment of the building's second floor to include the space and equipment needed to study Ebola survivors. Outside his office and stretching up the corridor, men and women sat in chairs that lined the walls, waiting to be seen by medical staff.

Since the survivor study was launched in Liberia in June 2015, more than 1,000 of the country's confirmed 1,500 Ebola survivors have agreed to take part. Their health will be monitored at semiannual checkups for five years. Each survivor is asked to bring four friends or relatives to one of the study's three sites. These are people with whom the patients have close contact but who were not infected with Ebola. Fallah says he hopes to enroll 6,000 close contacts who will serve as controls, helping researchers separate the health problems that are part of post-Ebola syndrome from those experienced by the general population in Liberia.

When Fallah presented the first findings from the study in February 2016, he had grim numbers: 60 percent of the approximately 1,000 virus survivors in the study reported eye problems, 53 percent said they suffered muscle aches and joint pain, and 68 percent reported neurological problems. His team dug deeper to learn what kind of damage the virus can inflict on the nervous system. At a meeting of neurologists in April 2016, they reported that nearly three quarters of Ebola survivors experienced headaches, 72 percent had depression, and more than half suffered memory loss and difficulty walking.

One in every four or five survivors had changes to the eye affecting vision. When Fallah's team looked more closely at those survivors, they found 10 percent had uveitis, a swelling of the middle layer of tissue in the eye wall.

The vision problems drew his attention early in the research. "We saw as the war went on—I mean, as the epidemic went on—that there were different manifestations among survivors, and that would drive us to do more in-depth substudies," he says. "And it was clear that the first target should be the eye."

Fallah looked to previous studies of Ebola survivors dating back to the 1990s and found that many described eye problems in the convalescent phase. After an outbreak in the Democratic Republic of the Congo in 1995, doctors examined 20 survivors, some of them more than two months postinfection. Four were

Illustration by Shizuka Aoki

© 2018 Scientific American

found to have eye pain, sensitivity to light, loss of visual acuity and uveitis. After another outbreak, this one in Uganda in 2007, 49 survivors were followed for more than two years. Along with memory loss, joint pain, sleep disorders and hearing loss, these people reported blurred vision and pain behind the eyes.

More recently, a study of eight patients who were treated for Ebola in U.S. hospitals found that all manifested various symptoms of post-Ebola syndrome up to four months after leaving the hospital. Six had psychological problems, including depression, anxiety and memory loss, and five suffered eye problems, including blurred vision and pain. There was no doubt the syndrome was real. But the existing data offered little explanation for how the virus can cause these problems.

AFTERSHOCK: Josephine Karwah, who survived Ebola infection only to be afflicted by other symptoms, stands outside a store in her village of Smell No Taste, Liberia.

UNDERSTANDING THE DAMAGE

THIS KIND OF CONFUSION has happened before, with another virus: HIV. Back in the 1980s when researchers were puzzled by this new viral threat, they tried to understand its effects by applying what they knew about other diseases. The same process is happening with Ebola, says Avindra Nath, a neurologist and scientist at the NIH who works closely with Fallah.

Nath has spent the better part of three decades studying infections of the brain. Although Ebola is not a retrovirus like HIV, Nath believes that years of research invested in studying HIV and the body's response to the infection have jump-started our understanding of how Ebola affects the nervous system. "Ebola has benefited from HIV research. A lot of us involved with Ebola made our careers with HIV, so we are quickly adapting our knowledge and techniques to studying these patients," he says.

Nath wonders if the neurological symptoms in Ebola survivors are a direct result of the virus or, instead, are triggered by the immune system's response to the infection. HIV, for instance, infects immune cells called macrophages in the brain, prompting the release of cytokines, small proteins that are toxic to nerve cells. Studies in monkeys have shown that Ebola infects macrophages and can set off a massive "cytokine storm." (Cytokines, chemical messengers between cells, trigger inflammation.) That can bring on hemorrhaging throughout the body, including the brain, which could explain the memory problems, headaches and movement disorders Nath has seen in Ebola survivors.

As the neurologist looks to HIV for clues to how Ebola affects the brain, others turn to different viruses to understand another symptom: the extreme fatigue reported by Ebola survivors. Studies have shown that up to a quarter of patients with dengue fever and close to 40 percent of Epstein-Barr patients develop fatigue after the acute illness. Inflammatory cytokines may be to blame. They can act on receptors in the brain to induce postinfection fatigue and loss of appetite.

Painful joints seem to be one of the more common symptoms of post-Ebola syndrome. In a study of survivors of the 1995 Congo outbreak, almost two thirds experienced joint pain two years after infection, and one third of a Ugandan outbreak's survivors suffered from joint pain two years later.

Lumps of immune system proteins that sit inside a joint like the hip or shoulder could cause irritation and swelling. Other components of the immune system, including antibodies, could explain or even act as a surrogate marker for joint pain. After the Congo outbreak, survivors who complained of painful joints were found to have higher antibody levels as compared with survivors who did not report joint pain. Another protein might be at work in causing joint pain, too: D-dimers, small chunks of protein that break off from blood clots, have been linked to joint pain in people recovering from other infections. Patients reporting joint pain after infection with the bacterium *Neisseria meningitidis* had high levels of D-dimers in their blood. Studies looking for D-dimer-level changes have not been done on Ebola survivors.

HIDING PLACES

AS FOR THE EYE DISEASE seen in many Ebola survivors, experts say it, too, could be a result of the immune response to the virus. Or, more ominously, the virus could be replicating in the eye long after it has been cleared from the blood. The eyeball offers a safe place for the virus to hide out, away from detection and interference by the immune system. In one survivor the eyeball was found teeming with Ebola. In September 2014 an American physician, Ian Crozier, fell sick with Ebola while working in Sierra Leone. Less than two months after he was discharged from a U.S. hospital, he felt pain in his left eye and noticed that its color had changed from blue to green. When doctors inserted a needle into Crozier's eye, they found more copies of the virus in his eyeball than had been in his blood when he was close to death weeks earlier.

The eyeball is not the only hiding place for Ebola. The testes, central nervous system and joint cartilage can act as sanctuary sites for a number of pathogens, including HIV. These vital structures are at risk of collateral damage when the immune system wages war on foreign invaders. So to protect themselves from the inflammatory response, they have adopted such clever mechanisms as immune-suppressing molecules and physical

Photograph by Seema Yasmin

© 2018 Scientific American

barriers. These protective measures make them great hiding spots for viruses. Hidden reservoirs could explain how Pauline Cafferkey, a Scottish nurse who recovered from Ebola, fell sick nine months after her blood tested negative for the virus and again a year after she was first infected.

If the testes harbors Ebola, that could explain why the virus persists in the semen of some survivors for months, even after they are free of symptoms. At the beginning of the West African outbreak, the World Health Organization cautioned people to practice safe sex for at least three months after their blood tested negative for Ebola. That advice was based on the 1995 Congo episode where the virus was found in the semen of survivors 82 days after the onset of symptoms.

But during the West African epidemic, Ebola virus lived in the semen of some survivors for a much longer time, more than a year after acute infection. At the conference in Boston, Fallah reinforced these findings, saying the virus was found in the semen of Liberian Ebola survivors 18 months after infection. In some men, the virus disappeared from the semen and then reappeared over the course of the year. (The WHO now advises male Ebola patients to practice safe sex for a year and to get their semen tested repeatedly.)

In his Monrovia office, Fallah has a patient file that belongs to a woman whose son died of Ebola in November 2015. The family reported no contact with anyone sick with Ebola or any survivors, but Fallah believes otherwise. He thinks the mother may have had sex with a survivor, not realized that she was sick with Ebola and passed the infection to her son.

Fallah had previously investigated a case of Ebola that was most likely transmitted via sex. In March 2015 a woman who died from Ebola was found to have had sex with a man who had been discharged from an Ebola treatment unit six months earlier. Blood samples from the man tested negative for Ebola, but a semen sample tested positive.

Fallah furrows his brow when talking about the woman who contracted Ebola from a survivor. That the virus can persist after many symptoms stop—even after a patient's blood appears clear—makes him anxious for two reasons: if Ebola hides out in people who seem healthy, only to reappear from compartments deep within the body to make them sick and potentially contagious, it could spark more outbreaks.

But finding the viral genome or bits of viral RNA in the bodily fluids of survivors does not prove they are contagious, he adds. What really worries Fallah is the stigma that these new findings place on survivors. "It's bad enough with post-Ebola syndrome that they have these symptoms we can't explain—and for who knows how long," he says. "Survivors are going through enough. Now imagine people are scared of them for fear of catching the virus."

TRAGEDY IN THE WOMB

A FEW DAYS AFTER Josephine left the Ebola treatment unit in Monrovia, while she was sleeping in her bed in Smell No Taste, she woke just after midnight. This time it was not nightmares or headaches; it was cramping in her abdomen. She rose to use the bathroom, and when she wiped herself she saw blood on the tissue. Then her water broke. "Ophelia!" she called for her older sister. They phoned for an ambulance but were told none were available. So they called a radio station in Monrovia for help. No one came.

Josephine paced up and down her bedroom, stopping to press her palms against the wall when it felt like her stomach was tearing. At 5 A.M., she wrapped herself in a maroon lapa, a traditional Liberian saronglike fabric, and staggered out of the house. If help would not come to her, she would find help on the streets. The village was asleep, sunrise still an hour away. Josephine walked alongside her house, clutching the walls to steady herself. As she screamed, women came out of their houses. "Help me, please help me," she cried. But no one would come near her, fearful of touching the woman who had left the Ebola treatment unit only a few days ago. When she reached the house at the corner of the dirt road, Josephine could no longer walk. She fell to the ground, her back against the wall and felt the baby between her legs.

Five women approached, unwrapping their lapas as they walked. They formed a semicircle around her so the male onlookers could not watch her give birth. Josephine pushed and screamed, and Miracle was born. What a chubby boy, she thought, lifting the silent child to her chest. But Miracle was not breathing.

No one would touch Josephine. The women stared as she rocked her baby and sobbed into her chest. Only her brother came close to her. He took Miracle from her arms and wrapped the baby and placenta in a yellow towel, ready for burial.

Josephine's mother had been a midwife before she died of Ebola. "Why isn't she here to help me now?" Josephine lamented. In the weeks that followed, there were more questions: Did Ebola kill Miracle, or was it because nobody would help? Would the baby have lived if an ambulance had come? Was the virus still lurking in her body, and would it harm any future pregnancies?

On visits to the Kennedy Medical Center for her survivor study appointments, Josephine asks Fallah these same questions. One afternoon she sits in his office wearing a leopard-print shirt and matching head wrap, waiting for his response.

Fallah worries the uterus may be another sanctuary site for Ebola, offering the virus a safe place to hide and still affect the body. Perhaps it could reappear and infect others. Then he wonders if the stress of being an Ebola survivor can cause a woman to give birth to a stillborn baby in the street with people watching but no one helping. He thinks, "When you can no longer sell soap in the market, when you have to wrap your money in tissue to buy vegetables, when your boyfriend stops loving you because you are an Ebola survivor—what impact does that have on a person's body? What could that do to the unborn child?"

This is what goes through his mind, but when Josephine asks, he says: "I don't know, Josephine. We are trying to find out." SA

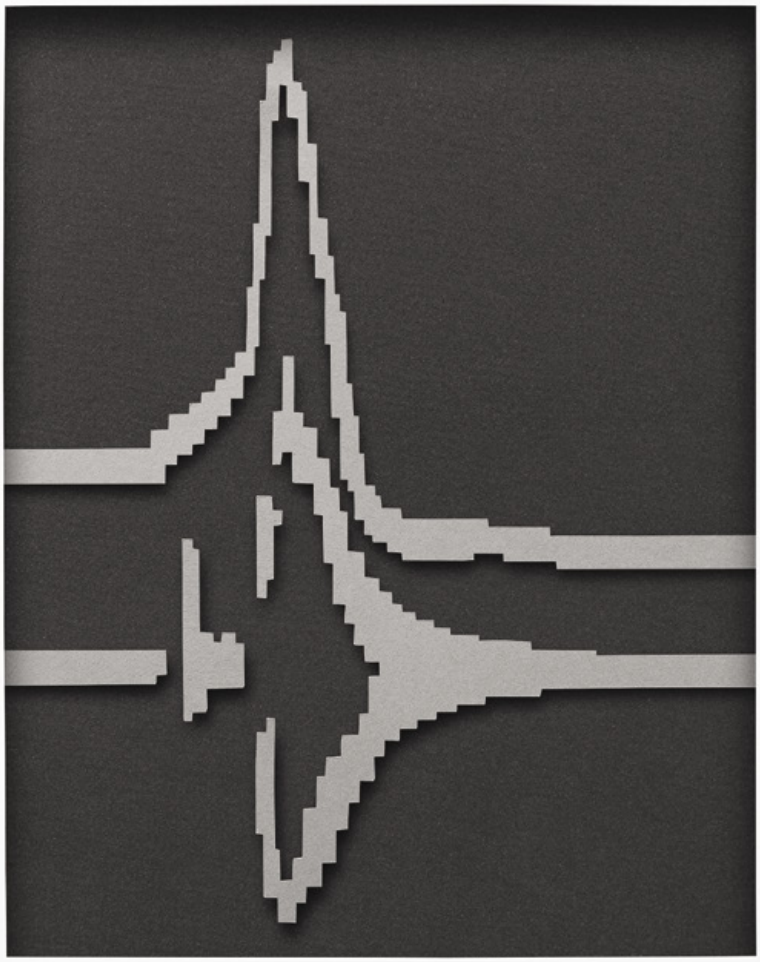
Seema Yasmin, a writer and medical doctor, reported for *Scientific American* from Liberia. Travel was supported by the Pulitzer Center on Crisis Reporting. Josephine Karwah's sister Salome died in childbirth the year after this article appeared.

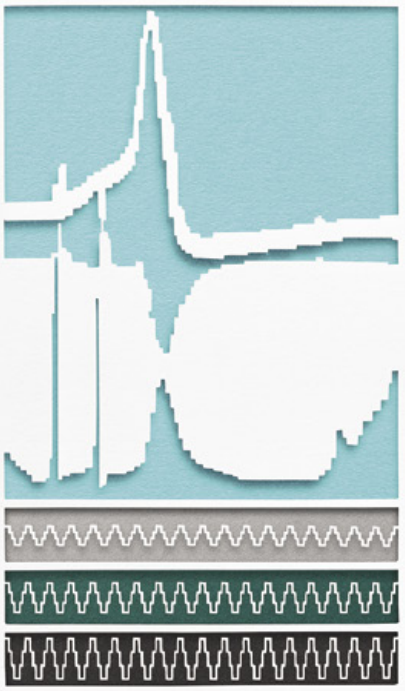
MORE TO EXPLORE

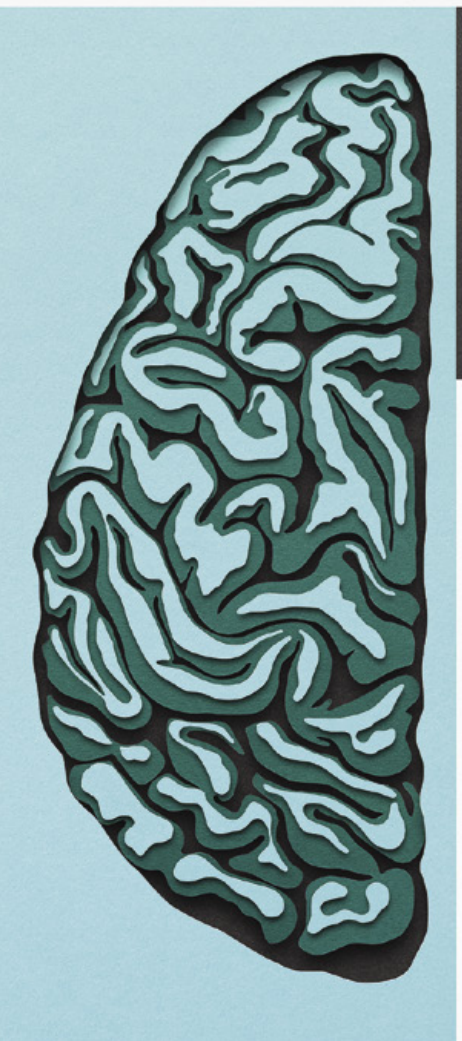
Possible Sexual Transmission of Ebola Virus—Liberia, 2015. A. Christie et al. in *Morbidity and Mortality Weekly Report,* Vol. 64, No. 17, pages 479–481; May 8, 2015.
Persistence of Ebola Virus in Ocular Fluid during Convalescence. Jay B. Varkey et al. in *New England Journal of Medicine,* Vol. 372, pages 2423–2427; June 18, 2015.
Serious and Common Sequelae after Ebola Virus Infection. Luke Hunt and Victoria Knott in *Lancet Infectious Diseases,* Vol. 16, No. 3, pages 270–271; March 2016.
A Woman Survives Ebola but Not Pregnancy in Africa. Seema Yasmin on ScientificAmerican.com; February 28, 2017.

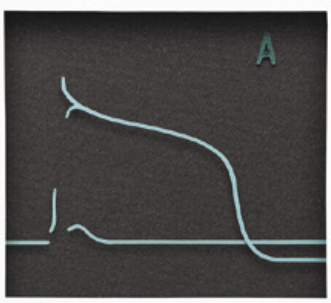
scientificamerican.com/magazine/sa

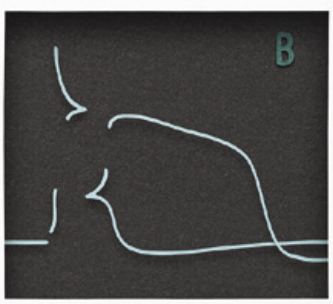
© 2018 Scientific American

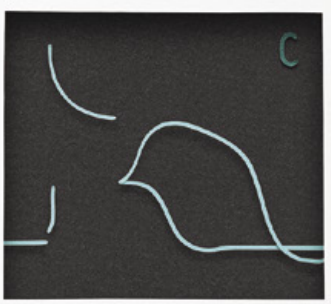


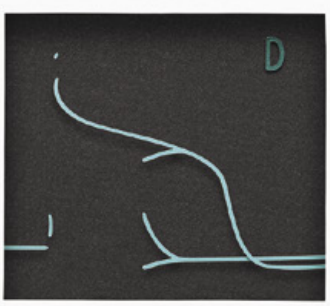


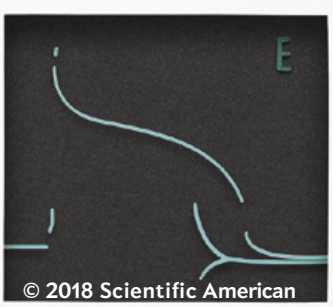


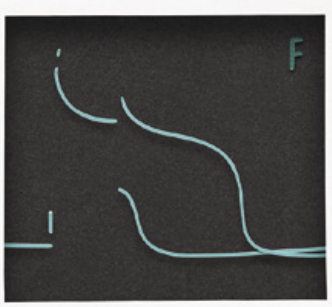












© 2018 Scientific American

THE BRAIN, Reimagined

Physicists who have revived experiments from 50 years ago say nerve cells communicate with mechanical pulses, not electric ones

By Douglas Fox

Illustration by Owen Gildersleeve

© 2018 Scientific American

A YOUNG WOMAN WITH WAVY BROWN HAIR AND MAROON NAILS lay on a gurney in a hospital room in Copenhagen. Her extended left arm was wired with electrodes. A pop pierced the air every few seconds—an electric shock. Each time, the woman's fingers twitched. She winced. She was to receive hundreds of shocks that day.

The woman, attended to by several physicians in laboratory coats, was renting out her arm for 1,000 Danish kroner, about $187. Thomas Heimburg, a physicist trained in quantum mechanics and biophysics, sat on a stool, safely out of the way, sketching on his iPad the details of a harsh experiment that he hoped would produce profound results.

The physicians had injected the woman's arm with the anesthetic lidocaine—a dose strong enough to deaden her limb for surgery. At first, the nerves in her arm did not respond to the shocks. But the attendants gradually dialed up the current. At this moment, the jolts were 40 milliamperes, nearly 10 times their original strength—similar to the electricity coursing through a five-watt lightbulb.

Pop—another shock. The woman's hand twitched like a dying snake. Heimburg paid no notice as he stared at a computer monitor on the wall. A waveform depicting the electric signal in the arm muscle and nerve leaped across the screen in one large spike—evidence that the ever increasing shocks had started to overcome the anesthetic. The nerve was now firing as strongly as it did before the woman was anesthetized. Heimburg was pleased. "The things that are written in books," he said quietly, "they are in contradiction to this."

Heimburg, who works at the Niels Bohr Institute in Copenhagen, famous for physics research, hopes to contradict lots of things written in books. This experiment, which I witnessed in December 2011, was designed to investigate a long-standing medical mystery.

Physicians have administered general anesthetics for 170 years. They have discovered dozens of effective compounds. When given at progressively higher doses, the drugs all silence nerve functions in the body and brain in the same distinct order: first memory formation, then pain sensation, then consciousness, and eventually breathing. This same sequence happens across all animals, from humans to flies.

Yet no one knows how anesthesia actually works. The molecular structures of nitrous oxide, ether, sevoflurane and xenon are so different that it is unlikely they exert their common effects by binding to equivalent proteins in cells, as other drugs do.

Heimburg thinks anesthetics work in a radically different way: by changing the mechanical properties of a nerve. If that is true, it means that nerve cells, or neurons, throughout the body and brain are mechanical machines, not the electric circuits scientists have believed in for decades. In Heimburg's view, the electric pulses are simply the side effects of a physical shock wave that ripples down the nerve, similar to the way sound waves travel. He thinks anesthetics silence nerves by soaking into the fatty membranes that encase nerve fibers, rendering them too soft to transmit the shock waves, like a guitar string too slack to twang.

It was tempting to dismiss Heimburg as nutty when I watched that experiment. But in the seven years since then, he and his colleagues have rolled out an array of evidence: delicate measurements of how mechanical waves move through single

IN BRIEF

Physicist Thomas Heimburg may upend the biology world. He says nerves do not fire electrically but mechanically. To prove it, he is reviving experiments done 50 years ago by a discounted neuroscientist.

Heimburg's radical idea is that a signal traveling down a pipelike nerve fiber is a compression wave, like sound, which temporarily changes the fatty membrane that coats the pipe from fluid to crystalline.

Biologists think Heimburg is revealing only side effects of an electric pulse, yet a few are acknowledging the two actions might work in concert, which would transform explanations about how the brain functions.

© 2018 Scientific American

nerve cells and of how much and how quickly the membranes can expand and contract, as well as studies showing how anesthetics alter these properties. Other scientists are starting to take an interest. Now Heimburg is preparing for a crucial experiment that could clinch his case: measuring the heat emitted by a single nerve cell as a pulse shoots through it.

Heimburg's work continues to demonstrate that the nerve pulse is more complex than most biologists may realize. The mechanical components may have been overlooked because of an accident of history: 50 years ago off-the-shelf instruments could readily measure the tiny electric impulses in neurons but not the mechanical ones. Hardware limitations influenced which discoveries scientists made and which ideas entered mainstream scientific thought. Heimburg's experiments are now reopening a decades-old scientific schism.

The story of the mechanical neuron holds lessons for all of science about biases and accidents of history. It also could change our basic understanding of nerves, brains and intelligence. Scientists have struggled to explain how brains achieve such daunting feats as face recognition and conversation while relying on proteins in neurons that are electrically noisy and unreliable. Heimburg is showing how the mechanical waves may compensate for this noise. If his theory proves out, he could rewrite biology. Or he might just be wrong.

HOT NERVES

THE NEURAL PULSE that scientists have tried for so long to explain lasts for only an instant. Step on a thumbtack, and your brain senses the pain within a fraction of a second. The signal travels through nerve fibers at up to 30 meters per second.

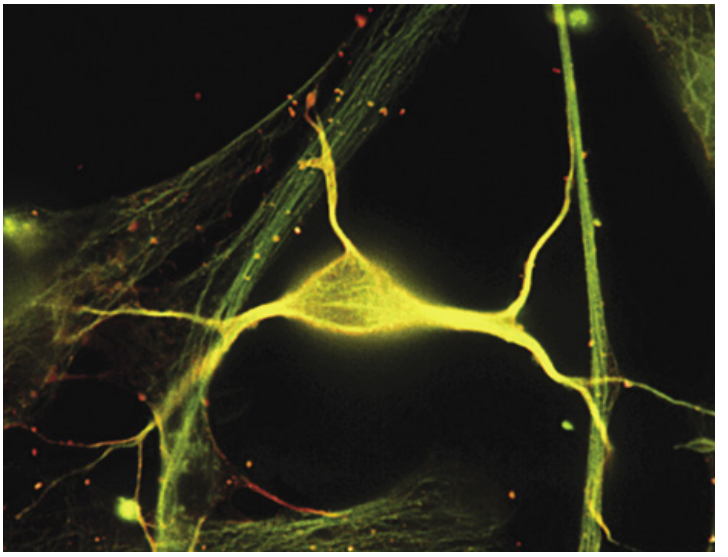
The fibers resemble tiny hollow pipes, finer than a hair. The pipe wall is formed by an oily cell membrane. Charged sodium and potassium atoms, called ions, hover around the inside and outside of the membrane. By the mid-1900s researchers had learned to stick electrodes into nerve cells to monitor the voltage across the membrane wall. They discovered that as a nerve pulse travels down the membrane and passes the electrode, the voltage spikes for several thousandths of a second. In 1952 two British scientists, Alan Hodgkin and Andrew Huxley, reported that the spike happens as sodium ions stream through the membrane wall from outside to inside. The voltage then reverses to normal as potassium ions gush through the membrane from inside to outside. The Hodgkin-Huxley model became the foundation of modern neurophysiology.

Hodgkin and Huxley received a Nobel Prize in 1963. But a few scientists continued to unearth observations that undermined their model, observations that Heimburg has re-created, even though some of those scientists had been written off as misguided.

Ichiji Tasaki, a senior neurobiologist at the National Institutes of Health for many years, was one of them. In 1979 he conducted an unorthodox experiment. Gazing through a microscope, he gingerly placed a fleck of shiny platinum atop a fine white thread—a nerve fiber bundle of a crab, laid bare by dissecting the animal's leg—and trained a laser onto the platinum. By measuring the reflection of the laser light, he could detect motions that would show whether the nerve bundle briefly widened or narrowed as an electric pulse passed by. He and his then postdoctoral fellow, Kunihiko Iwasa, took hundreds of measurements. After a week, the answer was clear: every time a pulse shot through the nerve fibers, they briefly widened, then narrowed again, within a few thousandths of a second.

The ripple was minuscule: the membrane surface rose by only about seven billionths of a meter. But it coincided perfectly with the passing electric pulse, confirming a suspicion Tasaki had harbored for years: that Hodgkin and Huxley were wrong.

As far back as the 1940s, researchers had noticed that as an electric pulse passes through a nerve fiber, the translucent cell briefly becomes more opaque. By 1968 Tasaki and another team found evidence suggesting that as the pulse arrives, molecules in the membrane physically rearrange themselves, then revert to their original configuration after the pulse passes.



NEURON, or nerve cell (*yellow*), in the brain's hippocampus—the center of long-term memory—is supported by proteins (*green* and *red*).

Then there was the heat. Researchers expected an electric pulse to release heat—common when electricity flows. But several teams discovered something strange. A nerve fiber's temperature rose several millionths of a degree Celsius as a pulse raced by, yet after it passed, the temperature quickly fell again. The heat had not dissipated; instead the nerve had reabsorbed most of it, also within a few thousandths of a second. For Tasaki, the transient widening, the rearranging molecules, and the heating and cooling pointed to a startling conclusion: the nerve signal was not just a voltage pulse; it was every bit as much a mechanical pulse. Scientists who listened to nerves with electrodes were missing much of the action.

Tasaki would spend the rest of his life probing these effects. He came to believe that they originated not in the cell membrane but in a layer of protein and carbohydrate filaments just underneath it. According to his theory, as the voltage pulse

ROBERT S. McNEIL *Science Source*

© 2018 Scientific American

arrives, the filaments absorb potassium ions and water—causing them to swell and warm—a process that then reverses itself after the pulse passes by.

As Tasaki pursued these ideas, he gradually fell out of step with the field. Other factors conspired against him. Having grown up in Japan, he spoke stilted English. "You [had] to know a lot of things to have a really substantive conversation with him," says Peter Basser, an NIH section head in neuroscience who knew Tasaki for 20 years. "And I think a lot of people thought he wasn't really as deep and perceptive as he was." And although Tasaki collaborated with visiting scientists, he did not produce student protégés who would carry his ideas forward.

Emblematic of the schism was the ideological rivalry that arose between Tasaki and another prominent NIH neuroscientist, Kenneth Cole, who adhered to the mainstream view. Although the two men occupied the same lab building from the 1950s to the 1970s, they barely spoke for 15 years, except at public presentations, where one would undermine the other by standing up in the audience and posing prickly questions.

Tasaki gave up his lab during an NIH reorganization in 1997 and moved into a small space in Basser's lab. He continued working seven days a week, well into his 90s. One day in December 2008, as he walked near his home, he lost his balance and banged his head on the ground. He died a week later at the age of 98.

By then, Tasaki's work had disappeared from sight. "I don't think anybody disputed that those things were being seen, because he was respected in the lab," said Adrian Parsegian, a biophysicist at the University of Massachusetts Amherst, who was at the NIH from 1967 to 2009. Rather Tasaki's findings "were explained away as not central" to nerve signaling—nothing more than side effects of the voltage pulse. The underlying scientific questions "didn't get resolved," he said. "One side got into the textbooks, and the other one didn't."

FATTY LIQUID BECOMES CRYSTAL

HEIMBURG CAME ACROSS Tasaki's work in the mid-1980s, while pursuing his Ph.D. at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany. Soon he found himself immersed in long sessions at the library, poring over old papers. He would eventually connect the dots in a different way than Tasaki had. He believed that the mechanical wave, the optical changes and the transient heat must occur in the fatty cell membrane of nerves throughout the body and brain, not in the protein and carbohydrate filaments below the membrane, as Tasaki had thought.

By the late 1990s Heimburg had begun doing his own experiments, compressing artificial cell membranes to see how they might respond to a mechanical shock wave. This work revealed something crucial: the membrane's oily lipid molecules are normally fluid and randomly oriented, but they hover close to what chemists call a phase transition. Squeeze the membrane just a little bit, and the lipids condense into a highly aligned liquid crystal.

These experiments led Heimburg to declare that a nerve pulse is a mechanical shock wave that travels down the nerve membrane. As it advances, it should squeeze the membrane's lipid molecules into a liquid crystal—a phase change that would release a small amount of heat, just as water does when it freezes. Then, as the tail end of the shock wave passes, a few thousandths of a second later, the membrane would revert to a fluid state, reabsorbing the heat. That brief transition into a liquid crystal and back would also cause the nerve membrane to widen briefly, just as Tasaki and Iwasa had seen when they shined a laser on that platinum fleck.

Heimburg's experiments went one crucial step further. They showed how the shock wave and phase transition might be linked to the voltage spike that occurs as the pulse passes by. Heimburg found that he could push a membrane into its liquid-crystal state simply by putting it under a voltage. "People applied voltage across biologic membranes for 70 years or so, and none of these electrophysiologists had ever checked" for a liquid-crystal structure, he said.

Textbook diagrams portray cell membranes as thin, passive sheets of insulation wrapped around pipelike nerve fibers. But physicists are starting to realize that cell membranes have surprising properties. They belong to a class of materials known as piezoelectrics, which can convert mechanical forces into electric forces, and vice versa. Quartz watches run on this principle. This means that a voltage pulse traveling down a membrane will carry with it a mechanical wave. And conversely, a mechanical wave traveling down a membrane will express itself as a voltage pulse.

When Heimburg and his fellow researcher Andrew D. Jackson first published the theory in 2005, they had still never observed one of these electromechanical pulses in motion. One of Heimburg's former students filled that gap. In 2009 Matthias Schneider, a biophysicist now at the Technical University of Dortmund in Germany, reported that he could trigger a mechanical wave by applying a voltage pulse to an artificial membrane. The pulse strength was similar to that found in nerve cells. The shock wave traveled at approximately 50 meters per second, similar to the speed at which thumbtack-triggered signals race from the foot to the brain. By 2012 Schneider had confirmed that the mechanical and voltage pulses were part of the same membrane wave.

Schneider's most important finding came in 2014, however. A key feature of a nerve pulse is that it is all-or-nothing. If a neuron receives a weak incoming shock, it will not fire a voltage pulse. If the shock is strong enough, it will fire. "There is a threshold," Schneider says. He found that the electromechanical waves on his artificial membranes were indeed all-or-nothing. The determining factor seemed to be whether the membrane was squished hard enough to force it into liquid-crystal form. Only then, he says, "you get a pulse."

ANESTHESIA EXPLAINED

WHY HAD HEIMBURG first committed to this view of nerves and anesthesia? Hoping to find out, I visited him at his office at the Niels Bohr Institute during the same week I witnessed the hospital experiment.

Heimburg had the bookshelves of a physicist, not a biologist, crammed with volumes by dead German physicists. Among them was a row of clothbound books by Hermann von Helmholtz, who in the mid-1800s formulated a key premise of thermodynamics, that energy can change form but cannot be created or destroyed. Helmholtz, incidentally, also measured the speed of nerve pulses. "I find it absolutely mandatory to read

© 2018 Scientific American

How Do Nerves Send Signals?

For decades scientists have endorsed a standard explanation for how nerve cells (neurons) transmit signals in the body and brain: each message is carried as an electric impulse that travels down a cell's long axon, jumping to the next neuron. But now a handful of physicists, who have performed exotic tests on the cells in action, say the signal is actually a mechanical pulse that ripples down the axon—akin to a sound wave or seismic wave. Some researchers say any physical pulse is just a side effect of the electric impulse. Settling these contentions could revise explanations of how the brain works.

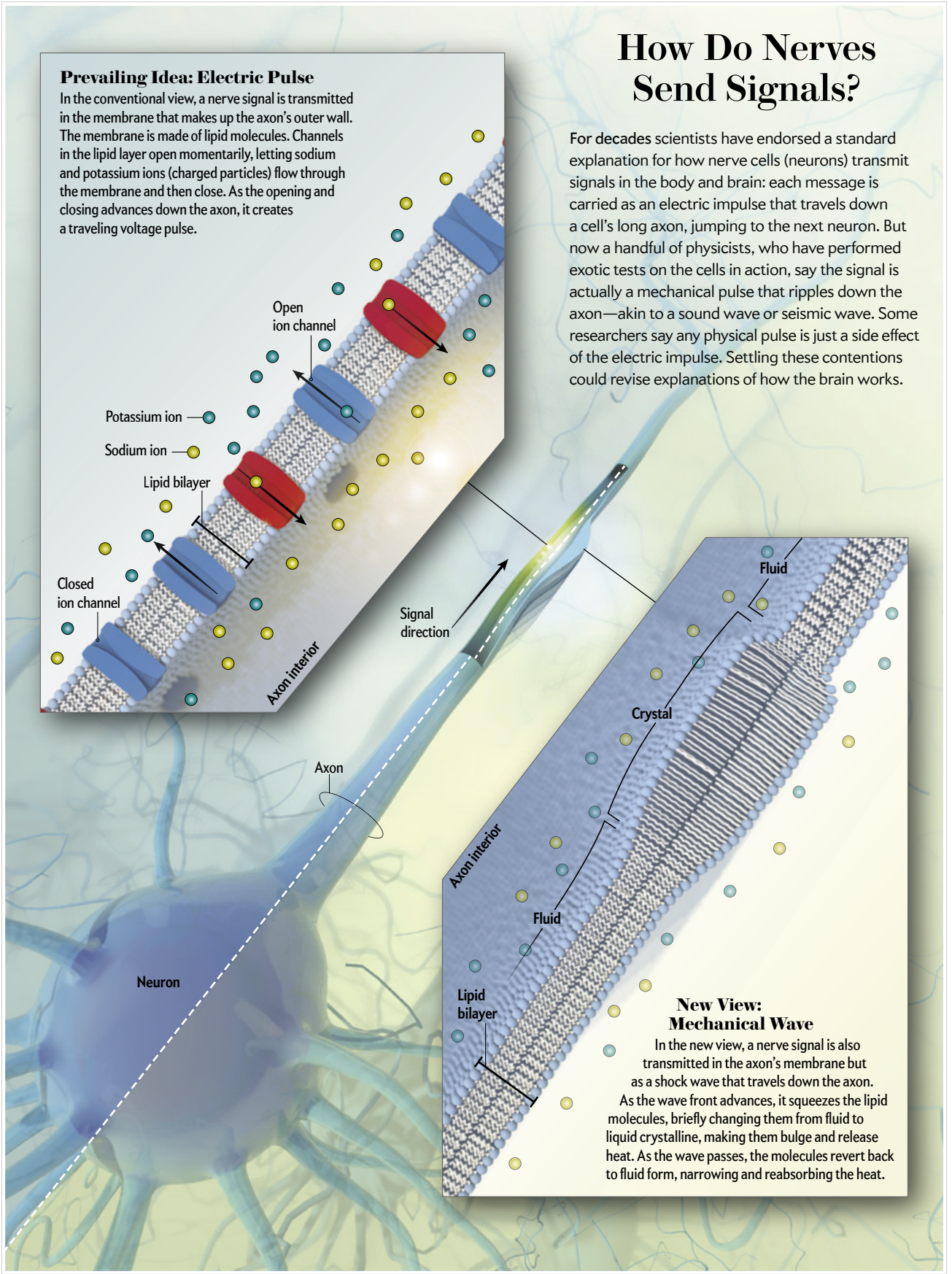

Prevailing Idea: Electric Pulse
In the conventional view, a nerve signal is transmitted in the membrane that makes up the axon's outer wall. The membrane is made of lipid molecules. Channels in the lipid layer open momentarily, letting sodium and potassium ions (charged particles) flow through the membrane and then close. As the opening and closing advances down the axon, it creates a traveling voltage pulse.
Open ion channel
Potassium ion
Sodium ion
Lipid bilayer
Closed ion channel
Axon interior
Signal direction
Axon
Neuron
Fluid
Crystal
Axon interior
Fluid
Lipid bilayer
New View: Mechanical Wave
In the new view, a nerve signal is also transmitted in the axon's membrane but as a shock wave that travels down the axon. As the wave front advances, it squeezes the lipid molecules, briefly changing them from fluid to liquid crystalline, making them bulge and release heat. As the wave passes, the molecules revert back to fluid form, narrowing and reabsorbing the heat.

Illustration by Falconieri Visuals

© 2018 Scientific American

these old texts," Heimburg said. They document the gradual discovery of fundamental connections among energy, temperature, pressure, voltage and phase transitions. These principles underlie Heimburg's ideas about nerve function, the ideas of a physicist pushing his way into another field. "Thermodynamics is the most profound science that we have," he said. "If you know thermodynamics, you are wise."

He was quick to point out weaknesses in popular explanations about anesthesia. Biologists think anesthetics silence nerves by binding to and thus blocking ion channels—valves in a nerve membrane that open and close to allow sodium or potassium ions to flow through. Biologists say the flow of ions propels voltage pulses down a nerve fiber—commonly portrayed as an electric signal. But because different anesthetics have vastly different molecular structures, Heimburg could not believe they could all bind to ion channels. That explanation was "completely ridiculous," he said, with a hint of frustration, as if pointing out something that should be obvious. Something "deeper, more profound," must be at work.

Heimburg's ideas were shaped in part by an old volume entitled *Studien über die Narkose,* or *Studies of Narcosis,* published by Ernest Overton in 1901. It recounts a particular experiment that caught his attention. Overton took dozens of different anesthetics and put each into a flask of water with a layer of olive oil floating on top. He shook each flask, then waited for the water and oil to separate again. He measured how much of each drug ended up in the oil versus the water. The more potent an anesthetic was in animals, the more strongly it moved into the oil, a striking result later confirmed for modern anesthetics. Olive oil and cell membranes are composed of the same oily molecules, called fatty acids. Heimburg surmised that the drugs might work by soaking into the cell membranes, altering their physical properties.

Experiments with synthetic membranes support that idea. When Heimburg infuses a membrane with an anesthetic, it prevents the membrane from becoming a liquid crystal. It does so by lowering the temperature (and raising the pressure) at which the phase transition from fluid lipid to crystalline lipid occurs—just as salt or sugar lowers the freezing point of water.

Heimburg reasoned that preventing this transition in a membrane would stop a mechanical pulse from advancing down a nerve fiber, explaining why anesthetics deaden nerves. And notably, he predicted it should be possible to overcome this effect. To create higher pressure to solidify a membrane using an electric shock, you have to crank up the current—exactly what the physicians did to the woman's arm at the hospital in Copenhagen. Stronger electric shocks did indeed overcome the anesthetic. If anesthesia can be overcome by pushing harder on a membrane with electricity, then it should also be reversible by increasing the physical pressure on a membrane.

Biologists demonstrated this way back in 1942. They used two different anesthetics, ethanol and urethane, to inebriate tadpoles to the point that they stopped swimming. Then the scientists put the animals in a hyperbaric chamber and raised the pressure to 136 times that of the atmosphere. The anesthetic effect vanished: the tadpoles resumed swimming. When the pressure was lowered, the animals again fell motionless. "It's very surprising," Heimburg said, with a smile. "How would you have the idea to put drunken tadpoles under pressure?"

THE EXISTENCE OF MECHANICAL WAVES IS NOT IN DOUBT, SAYS ONE NEUROSCIENTIST IN THE MIDDLE OF THE CONTROVERSY. "THE QUESTION IS WHETHER NEURONS ACTUALLY USE THEM TO DO SOMETHING USEFUL."

NO TOLERANCE FOR DEBATE

TO THIS DAY, Heimburg is frustrated by the way biologists react to his ideas, which he calls soliton theory (a soliton is a self-sustaining wave that maintains its shape as it travels). He has faced opposition from the moment he published his theory in 2005 in the *Proceedings of the National Academy of Sciences USA,* despite that journal's high regard.

One critic, Catherine Morris, a prominent neurobiologist emeritus at the Ottawa Hospital Research Institute, told me that the whole line of work reeks of superiority from a physicist who thinks he can simply march into a different field and set people straight. She summed this up in a favorite witticism of hers: "It strikes me as this business that physicists do, saying, 'We can approximate this cow as a single point.'"

To some extent, Morris's reaction is understandable. It is one thing to say that nerves are mechanical as well as electrical. It is quite another to reject the concept that ion channels play a role in nerve conduction—which Heimburg and Schneider do, in their biggest and most problematic departure from mainstream biology. Never mind that scientists have discovered hundreds of ion channel proteins. Or that the ion flows can be selectively altered with drugs. Or that mutations scientists can create in the proteins change the way neurons fire. "They just blithely ignore vast amounts of biology," says Morris, who spent 30 years studying ion channel proteins.

Heimburg and Schneider acknowledge that these proteins must serve some function. But they point to experiments, some by Heimburg, showing that ions can flow across artificial membranes even without channel proteins. They attribute this flow to transient holes that appear as the membrane shifts between fluid and liquid-crystal phases, and they think it happens in nerves in the body and brain.

Their skepticism reflects a cultural tendency in physics: a belief that all things should be explainable through thermodynamic principles. Biologists, they say, have neglected these principles as they fixate on proteins. A similar brand of puritanism may have facilitated the eventual dismissal of Tasaki's theory. He "did not like the term 'ion channels,'" said former postdoc Iwasa when we spoke in late 2017. This iconoclastic outlook may have guided Tasaki to discover things that others could not have, Iwasa said, "but later on, it may not have helped" him.

Brian Salzberg agrees. He studies nerve physics at the Uni-

© 2018 Scientific American

versity of Pennsylvania and began his neuroscience career in 1971, crossing paths occasionally with Tasaki. "He was a very clever experimenter, and I have no doubt that he measured real changes" in nerve thickness, Salzberg said earlier this year. "But he misinterpreted them." Salzberg says nerve fibers temporarily swell as a voltage pulse goes by in part because water molecules flow into the membrane through the same ion channels that let in sodium and then flow back out through the ion channels that let out potassium. If Tasaki had accepted the idea of ion channels, he might have been open to other interpretations of the mechanical wave.

But another powerful factor may have helped push Tasaki out of sight—holding an important lesson for all of science today.

IDEOLOGUES

IT IS INTRIGUING that the thermal energy of a firing nerve may be twice as large as the energy in the electric signal that has dominated neuroscience. The fact that these nonelectric features fell out of favor may stem, in part, from a quirk of history.

Tasaki was a gifted instrument builder who cut his scientific teeth in Tokyo during World War II. Faced with severe equipment shortages, he assembled his own instruments from stray electric components. Years later in the U.S., he used these skills to build exquisite, one-off instruments that measured the heat, or temporary expansion, of nerve cells.

Those devices, and expertise, never found their way to other scientists. Measuring the electric nerve signal was different. Scientists created easily transferable methods, such as inserting a tiny electrode into a cell membrane. As these techniques spread from one lab to another, so did the electrical view of nerve signaling. "There's a cultural bias," Parsegian admitted. "People look with a tool that they feel they understand, and they don't use one that they don't understand. It could have tilted the thinking."

Today the technical gaps are starting to disappear. As I checked in with Heimburg between 2011 and 2018, he gradually repeated one old experiment after another, using modern technologies to clarify the surprising things that Tasaki and others first saw decades ago. In 2014 Heimburg redid the drunken-tadpole experiment, using synthetic membranes instead of animals: as he cranked the pressure up to 160 atmospheres, the impacts of anesthetics were reversed—except that this time, Heimburg could link the effect directly to phase changes in the membrane. In 2016 he used microscopy to precisely measure, in a single cell, the mechanical wave that Tasaki and Iwasa first documented in 1979.

Heimburg, now 58, is seeking funding for what could be the most critical experiment of all: measuring the heat as a nerve pulse, or action potential, passes by. Tasaki had measured heat from bundles of fibers, but Heimburg plans to use a microchip that will measure the heat blip of a single neuron. This experiment could address a key criticism of his theory: that a nerve membrane's brief phase change from liquid to crystal should release, and reabsorb, more heat than Tasaki ever saw. Heimburg contends that the old experiments systematically underestimated the heat; because they measured many neurons, the heat reabsorption after early pulses canceled out the heat releases of later pulses. "The true signal is probably much higher," he told me in late 2017. If his measurements bear out, they could bolster his claim that the membrane transmits a mechanical wave.

Perhaps most significantly, other scientists are stepping in—outsiders who are not polarized by the old, calcified disputes. Nongjian Tao, a biosensor engineer at Arizona State University, is using lasers to track mechanical pulses in single nerve cells—like Tasaki and Iwasa did, except that Tao reflects his light directly off the nerve rather than a tiny platinum mirror, making the measurement more sensitive. He hopes to monitor hundreds of individual neurons in nerve networks at once, with lasers sensing mechanical waves as they ripple to-and-fro. Such work could answer a key question. "The existence of these [mechanical] effects is not in doubt," says Simon Laughlin, a neuroscientist at the University of Cambridge. "The question is whether neurons actually use them to do something useful."

Laughlin does not work on mechanical waves, but as someone who has studied ion channels for 45 years, he imagines that the waves could influence the little protein valves. Recent experiments show that the valves are extremely sensitive to mechanical forces in the membrane. If mechanical waves help to open and close ion channels, that could profoundly change our understanding of the brain because firing neurons mediate all thinking. Ion channels are notoriously noisy and jittery: even tiny thermal vibrations can cause them to pop open or close randomly. Information theorists have struggled for decades to explain how the brain can achieve reliable cognition using such unreliable channels. But mechanical waves could mean the openings and closings are purposeful. "That's a definite possibility," Laughlin says.

There are hints that this could be true. Some neurons in the mammalian cortex seem to violate the Hodgkin-Huxley theory. When they fire at high rates, their ion channels open more quickly, as a group, than expected. One explanation is that the channels are responding en masse to a sudden change in the membrane—the arrival of a mechanical wave that opens them more or less in unison, allowing them to fire faster than they otherwise could. The speed might allow them to transmit information at phenomenally quick rates—a possible basis for cognition. In this view, a nerve pulse is both electrical and mechanical.

Heimburg and Schneider occupy a strange place in all of this. They could perhaps one day share a Nobel Prize. Or they could end up nowhere, transfixed by the same insistence that gripped Tasaki for so many decades. The fact that some neuroscientists such as Laughlin and other experts such as Tao are interested in mechanical waves would seem like an important opening for the physicists. But Heimburg was steadfast when we spoke in early 2018. "What many people try to do is somehow rescue the Hodgkin-Huxley model by just combining it with the view that we have," he said. "But I personally … would not accept any kind of compromise between the two models." SA

Douglas Fox writes about neuroscience and extreme climates from California.

MORE TO EXPLORE

On Soliton Propagation in Biomembranes and Nerves. Thomas Heimburg and Andrew D. Jackson in *Proceedings of the National Academy of Sciences USA*, Vol. 102, No. 28, pages 9790–9795; July 12, 2005.

scientificamerican.com/magazine/sa

© 2018 Scientific American

MEMORY'S INTRICATE WEB

A technical revolution provides insight into how the brain links memories, a process critical for understanding and organizing the world around us

By Alcino J. Silva

© 2018 Scientific American

© 2018 Scientific American

Our memories

depend on our ability to recall details about the world—a child's face, a goose, a lake. To transform them into actual experiences, though, the brain must somehow merge these individual elements into an integrated whole—the look on that child's face when she sees a flock of geese suddenly take flight from a lakeside stand of reeds.

A cohesive sense of memory relies on other factors, too. Our survival over the millennia has depended on recalling not only the right information—say, a lion or a snake—but also the context. Did we encounter the animal during a surprise confrontation on an isolated stretch of African savanna or as part of an unhurried viewing at the San Diego Zoo?

To steer clear of other kinds of predators in our daily lives, we also need to be able to link memories over time: Judging whether a seemingly attractive investment is worth pursuing depends on the source of a recommendation—the probity, for instance, of the person who suggested it. Failing to connect the two can have disastrous consequences.

The field of neuroscience is starting to grapple with how the brain links memories across space and time. Until now, the vast majority of studies have focused on the way we acquire, store, recall and alter individual memories. Most memories, though, do not exist on their own as single, isolated entities. Instead one recollection summons the next, establishing intricate sequences of memories that help us to better predict and comprehend the world around us.

The fundamental mechanisms the brain uses to create these linked memories are beginning to reveal themselves—after 20 years of research in my laboratory and others. Understanding the physical processes involved in interweaving individual memories will do more than just provide insight into how the brain works. It may help us to prevent memory disorders that disrupt our ability to create and tie together memories.

A HAPPY ACCIDENT

WHEN WE BEGAN our studies of memory linking in the late 1990s, we lacked the tools and basic knowledge we needed to tackle this subject. A key first step in determining how memories are intertwined was our discovery of a concept called memory allocation, the realization that the brain uses specific rules to assign bits of learned information to discrete groups of neurons in regions of the brain involved in forming the memory.

Serendipity played a key role in the discovery of memory allocation. It started with a conversation I had with Michael Davis, a friend and colleague now at Emory University, during a visit to Yale University in 1998. Davis shared with me the findings of studies in which his lab manipulated a gene known as *CREB* to enhance emotional memory in rats—the association, for example, between a tone and an electric shock. Previously, my lab (now at the University of California, Los Angeles) and

IN BRIEF

Memory research has undergone a revolution: new technologies image the activity of individual neurons and even turn the cells on and off at precise moments, allowing brain scientists to perform experiments that were thought of as science fiction just a few years ago.

Techniques now available to neuroscientists have shown that memories are not randomly assigned to neurons in brain regions engaged in information processing and storage. Instead specific mechanisms determine which cells go on to store a given memory.

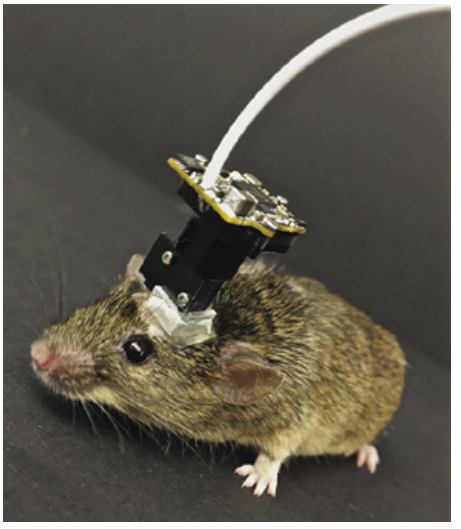
The brain's ability to control which neurons encode which memories is critical for strengthening memories and for connecting them, features that are disrupted in many neuropsychiatric disorders and during cognitive decline in aging.

PRECEDING PAGES: ANDREAS KUEHN *Getty Images (head)*; GETTY IMAGES (*web pattern*)

© 2018 Scientific American

other researchers had shown that the *CREB* gene was needed to form long-term memories. The *CREB* gene accomplishes this task by encoding a protein that regulates the expression of other genes needed for memory. During learning, some synapses (the cellular structures neurons use to communicate) are built up, or strengthened, so that they can facilitate interaction among cells. The CREB protein acts as a molecular architect of this process. Without its help, most experiences would be forgotten.

What surprised me was that Davis's group was able to improve memory, even though his lab increased CREB levels in only a subset of the overall population of neurons of the amygdala, a brain region critical for emotional memory. The question that lingered with me for months after my visit to Yale was, How did the memory end up in just the few cells where it could take advantage of the higher CREB levels? Could it be that CREB not only orchestrated memory formation but also helped to ensure that cells with CREB were more likely to be involved in memory formation? In our own investigations of CREB, we homed in on its function within specific brain regions we knew were involved with memory: the amygdala and the hippocampus. The latter stores an internal map of one's surroundings.



MICROSCOPE mounted on the head of a live mouse lets researchers inspect the activity of brain cells where memories are stored.

Science is just as much about finding questions as it is about answering them. What that conversation with Davis helped me realize is that neuroscientists knew very little about the rules, if, indeed, there were any, of how a given memory is allocated to the neurons in each of the brain regions that process and store our recollections. So we decided to look more closely.

Our first big break came after we recruited neuroscientist Sheena A. Josselyn, who had studied CREB in Davis's lab. In a series of animal experiments that she conducted in my lab and later with colleagues at her own lab at the University of Toronto, Josselyn used a virus to introduce extra copies of CREB into specific neurons within the mouse amygdala. She showed that those neurons were nearly four times more likely to store a fearful memory than neighboring ones.

In 2007, after almost a decade of effort, my lab, in collaboration with Josselyn's team, finally published evidence that emotional memories are not randomly assigned to neurons within the amygdala. Rather the cells tapped to store these memories are those that have more of the CREB protein. Just as important, subsequent experiments showed that CREB has a similar function in other brain regions, including the hippocampus and the cortex, the outermost layer.

SWITCHING MEMORIES ON AND OFF

TO CONFIRM CREB'S ROLE in memory allocation, we turned to newly developed methods that have transformed the study of memory in recent years. These lab techniques make it possible to either activate or switch off neurons—in effect, eliciting or silencing a memory.

As one example, Yu Zhou, then in my lab, genetically modified a small set of mouse amygdala neurons so that they had higher CREB levels and expressed another protein engineered by Edward Callaway's lab at the Salk Institute for Biological Studies in La Jolla, Calif. Callaway's nifty protein allowed us to silence the CREB neurons at a time of our choosing. When we shut off the neurons that had high CREB, leaving their counterparts with lower levels of the protein still active, emotional memory was suppressed, a result that provides evidence that neurons with higher levels of CREB are more likely to be involved in memory storage.

We knew that higher levels of CREB could determine which cells stored a given memory, but we did not know how this happened. Robert Malenka of Stanford University and his colleagues had discovered that increasing CREB in certain neurons meant they were more easily activated. Could this increase in excitability be the reason why neurons with higher CREB levels were picked for memory storage?

To address that question, Zhou modified amygdala neurons to produce more CREB. Using tiny microelectrodes, she determined how easily these neurons are activated, a measure of excitability. The results confirmed that the modified neurons were more easily switched on, compared with their unaltered counterparts. The elevated excitability (an enhanced readiness to receive and pass on electrical impulses that carry information between neurons) suggested that the cells may have been better prepared to initiate the set of processes needed for laying down a memory.

To test that idea, Zhou also looked at synaptic connections involving the neurons with more CREB. A considerable body of evidence has shown that increases in the strength of synaptic connections are critical for memory formation. After training the mice on a task that subsequently evoked emotional memories, she tested the strength of synaptic connections of the amygdala neurons with higher CREB levels to see whether they had stronger connections, compared with cells that had not been altered to produce more CREB.

To do this, she stimulated the synapses of these cells with a small electric current and recorded their responses with tiny electrodes embedded within the cells. As expected, the amygdala neurons with higher CREB had stronger synapses than other cells, a result consistent with the idea that they were more likely to have stored the emotional memory.

In still more recent work, Josselyn's lab demonstrated that a memory of a fearful experience could be stored in a predetermined set of amygdala neurons by genetically engineering

COURTESY OF DENISE J. CAI *Integrative Center for Learning and Memory, University of California, Los Angeles*

© 2018 Scientific American

them with a specific type of ion channel that increases the excitability of these neurons. Ion channels form pores on the surface of the cells, and the particular ones that Josselyn chose allowed these cells to be more easily activated. Similarly, neuroscientist Albert Lee's lab at the Howard Hughes Medical Institute's Janelia Research Campus in Ashburn, Va., reported that artificially increasing the excitability of hippocampal neurons in a specific place while animals ran around a track made those neurons more likely to respond to that location in the track, a result consistent with our findings that excitability has a critical role in determining which cells are engaged in storing a given memory.

Finally, our group, as well as Josselyn's, took advantage of a groundbreaking technology called optogenetics that uses light to either activate or inhibit neurons. We used the technique to switch on specific neurons that had higher CREB levels. Thomas Rogerson and Balaji Jayaprakash, both then in my lab, began by engineering amygdala neurons to produce more CREB and channelrhodopsin 2 (ChR2), an ion channel activated by blue light. We then showed that we could artificially trigger recall of a fear memory in mice when we used the light to turn on amygdala neurons with higher CREB but not ones with lesser levels of the protein, confirming that the memory was stored in those neurons.

THE FUNDAMENTAL MECHANISMS THE BRAIN USES TO CREATE THESE LINKED MEMORIES ARE BEGINNING TO REVEAL THEMSELVES.

LINKING UP

IN 2009 I WAS ASKED to write an article on our memory research, and I took that opportunity to introduce our ideas on how memories are linked over time. CREB's ability to regulate which cells form a given memory—in other words, memory allocation—led me to the hypothesis that this process may be key to the ability to connect separate memories, what my lab now calls the "allocate to link" hypothesis. Because memory allocation occurs in a subset of neurons having higher CREB that are more easily activated, this process primes these neurons to readily store another memory. When two memories share many of the same neurons, they are formally linked.

Consequently, activation of those neurons during recall of one of the two memories triggers recall of the other. Key to this idea was the prediction that two memories formed closer in time—both within a day—are more likely to be linked than when they are separated by longer periods. With intervals much longer than a day, the second memory no longer benefits from the excitability triggered by the first memory and so is stored in a different population of neurons. The time-limited nature of memory linking makes sense because events that occur within the span of a day are far more likely to be relevant to one another than those separated by, say, a week.

Writing the article and outlining these ideas drew me even more to the challenge of how we might test them. The allocate-to-link hypothesis was straightforward, but it was not at all clear how we would confirm its legitimacy. Testing had to wait for the right time.

The situation changed when Denise J. Cai and Justin Shobe, both then in my lab, joined the project. Cai came up with a clever idea. Together with Shobe, she exposed mice to two chambers during the same day within an interval of five hours, hoping that the memories of the two chambers would be linked. Later she gave them a mild paw shock in the second chamber. As expected, when she subsequently placed the mice in the chamber where they received the shock, they froze, presumably because they remembered that they had received a shock there. Mice freeze as a natural reaction to fear because most predators notice prey better when they move.

The critical result emerged when Cai and Shobe placed the mice in the neutral chamber. We reasoned that if the memories from both chambers were linked, the mice in the neutral space would be reminded of being shocked in the other chamber and thus would freeze in anticipation—and that is exactly what we found.

We also guessed that the two memories would be less likely to be linked if they were separated by a seven-day interval. And indeed, reexposing the animals to the neutral chamber after the longer time span did not remind them of the shock chamber, and they did not freeze. In general, with time intervals much longer than a day, memories remain unlinked. Remarkably, studies in Jill Leutgeb's lab at the University of California, San Diego, suggested that the very brain region in our experiments has the ability to distinguish between time intervals within the time frames we had tested (many hours), a result that was in agreement with our findings.

These behavioral findings were exciting, but they did not test a key prediction of the hypothesis—that discrete memories formed at closely spaced intervals are stored in the same brain area in overlapping populations of neurons. This physical overlap links the two memories, so that the recall of one brings to mind the other.

VISUALIZING MEMORIES

TO REALLY TEST the allocate-to-link hypothesis would require nothing short of being able to see memories in the brain as they were being created. Techniques for imaging neurons in live mice are already in use, but they all required that the heads of the mice be fixed to large microscopes, a setup not conducive to the behavioral experiments needed to test the hypothesis.

I find it amazing, though, how many times in my career the right technique has come along just when we need it the most. I happened to attend a seminar at U.C.L.A., given by Mark Schnitzer of Stanford, that described a tiny microscope his lab had just invented that could visualize the activity of neurons in freely moving mice. This two- to three-gram microscope can be mounted like a hat on an animal's head. The instrument was what our group needed to track the neurons activated by a given memory. Schnitzer and his colleagues had shown that it could be used to track the very neurons we were studying across time. Therefore, we could use their revolutionary miniature microscopes to test a key prediction of our allocate-to-link

© 2018 Scientific American

hypothesis: many of the neurons active during one memory become active again a few hours later during the creation of another linked memory.

We were so excited by the promise of this wonderful invention that we decided to engineer our own version of the microscope. We teamed up with Peyman Golshani's and Baljit Khakh's labs, both at U.C.L.A., and together we hired a talented postdoctoral fellow, Daniel Aharoni, who went on to engineer what we came to call the U.C.L.A. miniscopes. Similar to the Schnitzer microscopes, our miniscopes are equipped with a lens that could be embedded near the brain cells we wanted to record from. The device is snapped onto a base plate secured to the animal's skull, holding it stable during training tasks and memory testing. Just as we borrowed techniques from other researchers, we were also glad to share. We are avid supporters of the open-source movement in science and have made our designs and software for the U.C.L.A. miniscopes available to hundreds of other groups worldwide.

To visualize the activity of neurons with the miniscopes, Cai and her colleague Tristan Shuman took advantage of an imaging technique that genetically engineers neurons in an animal so that they fluoresce when calcium levels in the cells rise—it is known as a genetically encoded calcium indicator.

We decided to focus on the CA1 region of the hippocampus because of its role in learning and remembering places, such as the chambers that we had used in our behavioral experiments. The mice wearing their miniscope hats were placed in the two chambers. We wanted to know whether the time interval between exposures to the different chambers affected which neurons were activated.

The results were more than we had expected! Essentially our miniscope and behavioral experiments showed that when the mice linked the memories of the two chambers, many of the CA1 neurons that became active when the animals visited the first chamber were also switched on when they explored in the second chamber. If the interval between visits was about five hours, the mice formed two memories in a similar cluster of neurons. When the time lapse increased to seven days, this overlapping pattern of activation did not appear.

We were delighted by this finding because it confirmed a basic premise of the allocate-to-link hypothesis: memories couple when they are stored in overlapping populations of neurons. If you later reactivate an ensemble of neurons formed for either of two memories, it stimulates the other one and facilitates its recall.

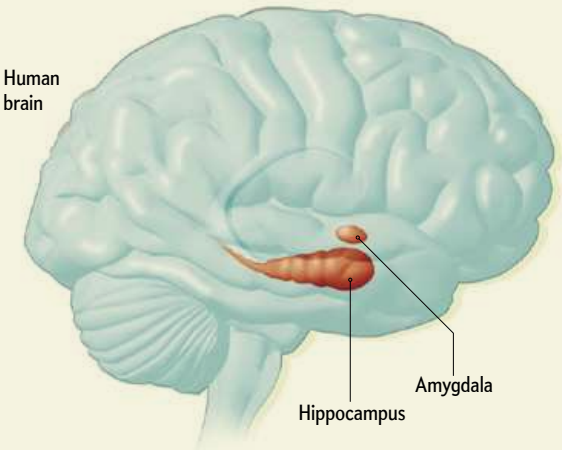
TAGGING MEMORIES

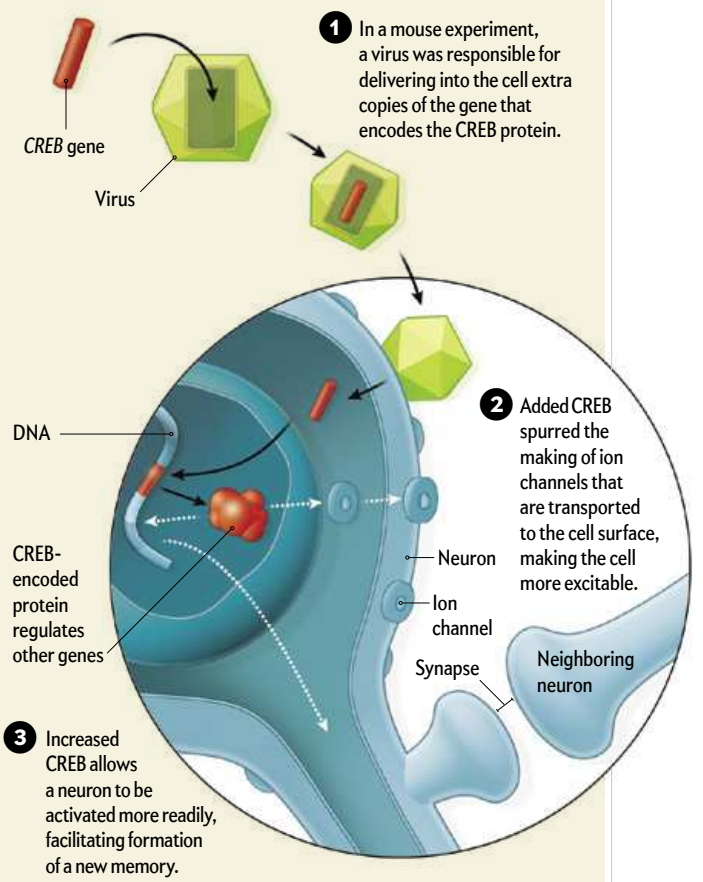
TO FURTHER VALIDATE the miniscope results, Cai turned to another method developed by neuroscientist Mark Mayford, now at the Scripps Research Institute. This experiment involved Mayford's technique, called the TetTag system (for *tet*racycline *tag*). When a memory is formed during a transgenic mouse's visit to a chamber, TetTag marks activated neurons with a fluorescent marker that remains intact for weeks.

Postmortem studies of the animals can then compare the recently activated neurons—tagged by genes that are expressed immediately after memory formation—with those marked by the long-lasting tag. This step identifies not only neurons switched on by one event—in which case a neuron

Memory Makers

Key brain regions play a role in forming memories. The amygdala is essential for memories with emotional content, and the hippocampus is involved in creating memories of experiences. My laboratory performed a mouse experiment that showed that cells in which my team increased the levels of a protein called CREB were more likely to encode a memory. —*A.J.S.*



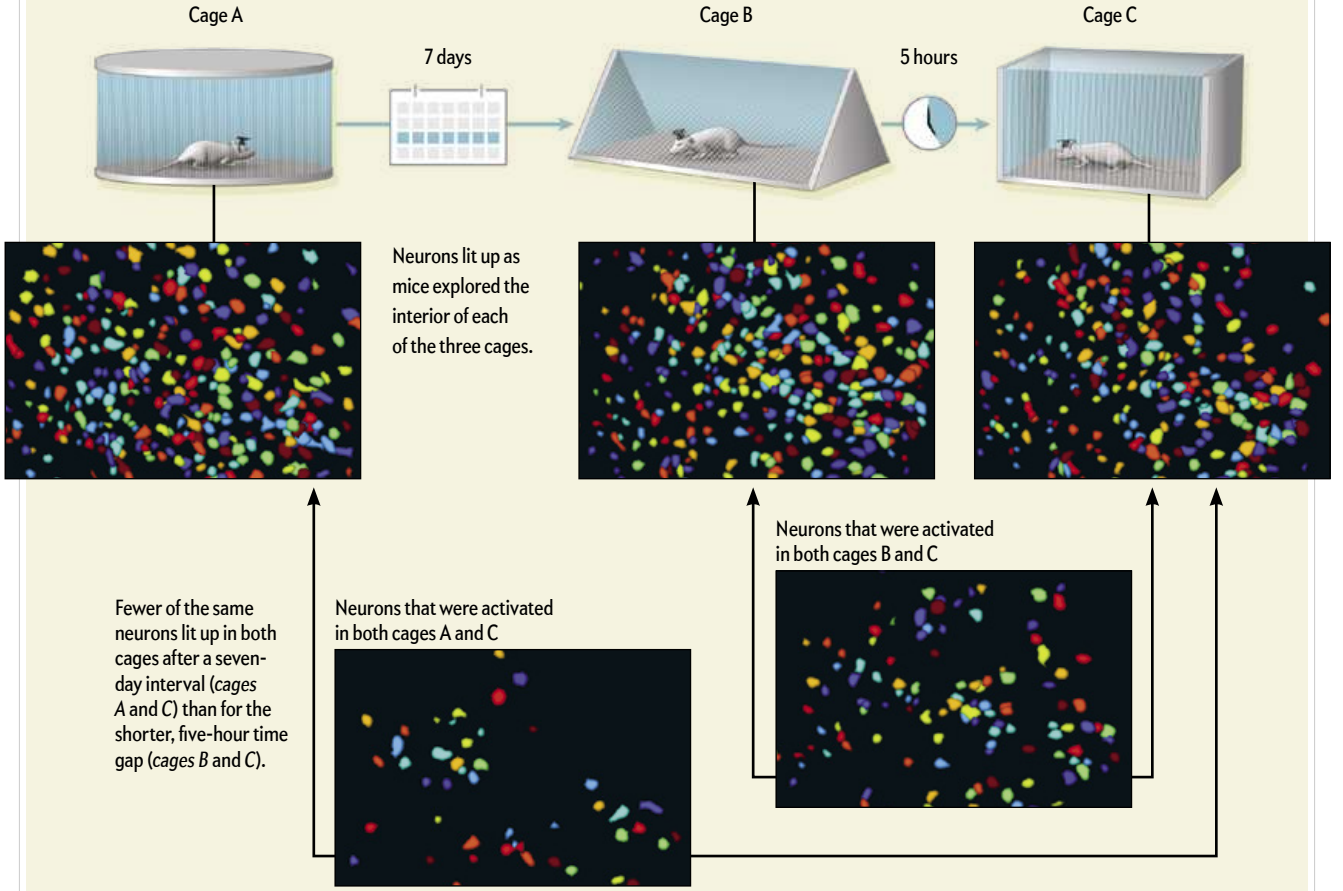

CREB gene
Virus
1 In a mouse experiment, a virus was responsible for delivering into the cell extra copies of the gene that encodes the CREB protein.
DNA
2 Added CREB spurred the making of ion channels that are transported to the cell surface, making the cell more excitable.
CREB-encoded protein regulates other genes
Neuron
Ion channel
Synapse
Neighboring neuron
3 Increased CREB allows a neuron to be activated more readily, facilitating formation of a new memory.

Illustrations by Tami Tolpa

© 2018 Scientific American

Remembrance of (Linked) Things Past

The "Proustian moment"—when one recollection leads to the next—has now gained a solid footing in the brain sciences. Experiments have shown that a mouse exposed to two chambers—say, B and C—links the two together in its memory if exposed to the two enclosures at an interval of five hours. But a mouse does not remember cages A and C together if the time period is separated by seven days. The linked recall of cages B and C takes place because many of the same neurons used to store the memories of the two cages turn on at the same time, unlike those for cages A and C.


Cage A
7 days
Cage B
5 hours
Cage C
Neurons lit up as mice explored the interior of each of the three cages.
Neurons that were activated in both cages B and C
Fewer of the same neurons lit up in both cages after a seven-day interval (*cages A and C*) than for the shorter, five-hour time gap (*cages B and C*).
Neurons that were activated in both cages A and C

features a single fluorescing tag—but also those activated by two occurrences: the glowing of both tags.

Using the same experimental setup as before, Cai and her team showed that during a short, five-hour interval, the overlap between the neurons encoding each of two memories with double tags was significantly greater than would be expected by chance. For a seven-day interval, the overlap between two experiences was not significantly above the level of chance.

Other experiments by Josselyn's Toronto team provided still more evidence of the validity of our memory-linking hypothesis. Not only did her group carry out a different version of the neuronal tagging experiment, the scientists also found independent behavioral evidence for memory linking. The Toronto researchers reasoned that if populations of neurons encoding two memories overlapped, increases in CREB levels triggered by the first memory would also strengthen a second memory. But instead of exposing the mice to different places, as in our work, Josselyn's team trained the animals to learn to recognize two different tones. Training on the first tone strengthened the memory of a second tone if the two training sessions occurred within six hours but not from six to 24 hours.

Recently Kaoru Inokuchi and his colleagues at the University of Toyama in Japan took this analysis a step further. They used optogenetics to inactivate the group of cells that was shared by two different emotional memories while leaving other cells undisturbed, including those that were unique to each of the two memories. The investigators showed that by inactivating the shared cells, they were able to disrupt the linking between the two memories without affecting recall of each individual memory. This elegant experiment provided direct evidence that the neurons shared by two memories are key for memory linking. It also added to the number of labs

FROM "A SHARED NEURAL ENSEMBLE LINKS DISTINCT CONTEXTUAL MEMORIES ENCODED CLOSE IN TIME," BY DENISE J. CAI ET AL., IN *NATURE*, VOL. 534; JUNE 2, 2016 *(neuron images)*

© 2018 Scientific American

that provided independent evidence for our fledgling allocate-to-link hypothesis.

IMPROVING MEMORY IN AGING

NEXT, WE DECIDED TO STUDY memory linking in older mice. Compared with young mice, older mice have lower levels of CREB in the brain, including in neurons in the CA1 area of the hippocampus, and consequently lower excitability. Knowing that, we predicted that aging mice should run into difficulties in linking memories. So Cai and her colleagues set about repeating many of the same experiments we had already completed in older animals. The results surprised us. Experienced scientists know that hypotheses are just tools. We do not expect them to be necessarily right. Inevitable failures help us to reshape our ideas along the way. But this time, our hunches proved correct.

WHEN TWO MEMORIES SHARE MANY OF THE SAME NEURONS, THEY ARE FORMALLY LINKED.

I still remember when Cai burst into my office, slightly out of breath. She told me that the middle-aged mice, despite remembering each individual chamber, indeed had problems linking the memories, even when they were exposed to them five hours apart, an interval that presented no difficulty for younger mice. Compared with the young adult mice, the miniscope imaging of the older animals revealed a lack of overlap between stored memories.

We were both excited but also skeptical, so we went right back and repeated the experiments. The second time around, the results became only more convincing. The neurons in middle-aged mice with lower CREB levels did not link memories as easily as those in young mice.

These results emboldened us to broaden the scope of our investigation. Could we increase artificially the excitability of a subset of CA1 neurons just when the older mice explored the two chambers, ensuring that some of the CA1 neurons activated in one chamber were also switched on when the animals moved to the second?

To accomplish this, we took advantage of a groundbreaking method that genetically engineers receptors onto the surface of a cell, which allows control over the cell's function. The technique bears the memorable techie acronym DREADD (for *d*esigner *r*eceptors *e*xclusively *a*ctivated by *d*esigner *d*rugs). Activating the DREADD receptors allowed us to turn on the same subset of CA1 neurons while the animals explored both chambers, thereby forging a link between their memories of the two enclosures.

I must confess that at first the idea for this experiment sounded preposterous. There are any number of reasons why it could have failed. For one thing, memories for places involve many millions of neurons spread throughout multiple interconnected brain regions, not just the CA1 region. Aging could have affected memory-linking processes in many, if not all, of these areas. Thus, even if we were successful in increasing excitability in a subset of CA1 neurons, these cells may not have been the right ones. What is more, we may not have triggered the right levels of excitability.

But the experiment worked. The key for this type of Hail Mary trial is to balance investment in time and money with the potential payoff that may be forthcoming. Nevertheless, in this case, I can safely say that luck was on our side. By restoring increases in excitability in a specific subset of CA1 neurons of middle-aged mice, we were able to allocate the two memories to many of the same CA1 neurons and thus restore memory linking in these middle-aged mice.

Research from other labs in both rodents and humans has also elucidated how one memory can be intertwined with another. The late neuroscientist Howard Eichenbaum of Boston University demonstrated that rats are capable of finding connections between memories that share content. Neuroscientist Alison Preston of the University of Texas at Austin and her colleagues showed that when memories share content, humans can link them more easily. Recalling one will likely bring back the other. In 2017 the same group also showed that temporal proximity facilitates the integration of memories across time, a result in humans with striking parallels to our results in mice.

The growing arsenal of tools at our disposal to measure and control neural activity is beginning to unravel the mechanisms our brain uses to organize information. Our team is now trying to extend this work in new ways. Together with computational neuroscientist Panayiota Poirazi of the Institute of Molecular Biology and Biotechnology at the Foundation for Research and Technology–Hellas in Greece, we are building computer models to simulate how and when memories link up. We are also trying to figure out the mechanisms that control the time intervals needed for memory linking in different brain structures.

So far a number of broad-ranging experiments carried out by multiple labs all strongly support the allocate-to-link hypothesis. We hope that an understanding of how memories become entangled may help us to develop treatments for memory problems that are common across a wide swath of psychiatric conditions, from age-related cognitive decline to schizophrenia, depression and bipolar disorder. Beyond clinical implications, the studies we have described reflect an exciting new era in memory research in which the experiments we do are no longer limited by the techniques we have at our disposal but by the reaches of our imagination. SA

Alcino J. Silva is a Distinguished Professor and director of the Integrative Center for Learning and Memory at the University of California, Los Angeles. His laboratory (www.silvalab.org) studies memory mechanisms, as well as the causes and treatments for memory disorders.

MORE TO EXPLORE

Synaptic Tagging during Memory Allocation. Thomas Rogerson et al. in *Nature Reviews Neuroscience,* Vol. 15, No. 3, pages 157–169; March 2014.

Finding the Engram. Sheena A. Josselyn et al. in *Nature Reviews Neuroscience,* Vol. 16, No. 9, pages 521–534; September 2015.

A Shared Neural Ensemble Links Distinct Contextual Memories Encoded Close in Time. Denise J. Cai et al. in *Nature,* Vol. 534, pages 115–118; June 2, 2016.

Competition between Engrams Influences Fear Memory Formation and Recall. Asim J. Rashid et al. in *Science,* Vol. 353, pages 383–387; July 22, 2016.

Temporal Proximity Promotes Integration of Overlapping Events. D. Zeithamova and A. R. Preston in *Journal of Cognitive Neuroscience,* Vol. 29, No. 8; pages 1311–1323; August 2017.

scientificamerican.com/magazine/sa

© 2018 Scientific American

© 2018 Scientific American

LANGUAGE IN A NEW KEY

Much of Noam Chomsky's revolution in linguistics—including its account of the way we learn languages—is being overturned

By Paul Ibbotson and Michael Tomasello

THE IDEA THAT WE HAVE BRAINS HARDWIRED WITH A MENTAL TEMPLATE for learning grammar—famously espoused by Noam Chomsky of the Massachusetts Institute of Technology—has dominated linguistics for almost half a century. Recently, though, cognitive scientists and linguists have abandoned Chomsky's "universal grammar" theory in droves because of new research examining many different languages—and the way young children learn to understand and speak the tongues of their communities. That work fails to support Chomsky's assertions. The research suggests a radically different view, in which learning of a child's first language does not rely on an innate grammar module. Instead the new research shows that young children use various types of thinking that may not be specific to language at all—such as the ability to classify the world into categories (people or objects, for instance) and to understand the relations among things. These capabilities, coupled with a

IN BRIEF

Noam Chomsky has been a towering giant in the field of linguistics for many decades, famed for his well-known theory of universal grammar.

Chomsky's idea of a brain wired with a mental template for grammar has been questioned, based on a lack of evidence from field studies of languages.

The theory has changed several times to account for exceptions that run counter to its original postulations—marking a retreat from its ambitious origins.

Alternatives to universal grammar posit that children learning language use general cognitive abilities and the reading of other people's intentions.

Illustration by Owen Gildersleeve

© 2018 Scientific American

unique human ability to grasp what others intend to communicate, allow language to happen. The new findings indicate that if researchers truly want to understand how children, and others, learn languages, they need to look outside of Chomsky's theory for guidance.

This conclusion is important because the study of language plays a central role in diverse disciplines—from poetry to artificial intelligence to linguistics itself; misguided methods lead to questionable results. Further, language is used by humans in ways no animal can match; if you understand what language *is,* you comprehend a little bit more about human nature.

Chomsky's first version of his theory, put forward in the mid-20th century, meshed with two emerging trends in Western intellectual life. First, he posited that the languages people use to communicate in everyday life behaved like mathematically based languages of the newly emerging field of computer science. His research looked for the underlying computational structure of language and proposed a set of procedures that would create "well-formed" sentences. The revolutionary idea was that a computerlike program could produce sentences real people thought were grammatical. That program could also purportedly explain the way people generated their sentences. This way of talking about language resonated with many scholars eager to embrace a computational approach to … well … everything.

As Chomsky was developing his computational theories, he was simultaneously proposing that they were rooted in human biology. In the second half of the 20th century, it was becoming ever clearer that our unique evolutionary history was responsible for many aspects of our unique human psychology, and so the theory resonated on that level as well. His universal grammar was put forward as an innate component of the human mind—and it promised to reveal the deep biological underpinnings of the world's 6,000-plus human languages. The most powerful, not to mention the most beautiful, ideas in science reveal hidden unity underneath surface diversity, and so this paradigm held immediate appeal.

But evidence has overtaken Chomsky's theory, which has been inching toward a slow death for years. It is dying so slowly because, as physicist Max Planck once noted, older scholars tend to hang on to the old ways: "Science progresses one funeral at a time."

IN THE BEGINNING

THE EARLIEST INCARNATIONS of universal grammar in the 1960s took the underlying structure of "standard average European" languages as their starting point—the ones spoken by most of the linguists working on them. Thus, the universal grammar program operated on chunks of language, such as noun phrases ("The nice dogs") and verb phrases ("like cats").

Fairly soon, however, linguistic comparisons among multiple languages began rolling in that did not fit with this neat schema. Some native Australian languages, such as Warlpiri, had grammatical elements scattered all over the sentence—noun and verb phrases that were not "neatly packaged" so that they could be plugged into Chomsky's universal grammar—and some sentences had no verb phrase at all.

These so-called outliers were difficult to reconcile with the universal grammar that was built on examples from European languages. Other exceptions to Chomsky's theory came from the study of "ergative" languages, such as Basque or Urdu, in which the way a sentence subject is used is very different from that in many European languages, again challenging the idea of a universal grammar.

These findings, along with theoretical linguistic work, led Chomsky and his followers to a wholesale revision of the notion of universal grammar during the 1980s. The new version of the theory, called principles and parameters, replaced a single universal grammar for all the world's languages with a set of "universal" principles governing the structure of language. These principles manifested themselves differently in each language. An analogy might be that we are all born with a basic set of tastes (sweet, sour, bitter, salty and umami) that interact with culture, history and geography to produce the present-day variations in world cuisine. The principles and parameters were a linguistic analogy to tastes. They interacted with culture (whether a child was learning Japanese or English) to produce today's variation in languages as well as defined the set of human languages that were possible.

Languages such as Spanish form fully grammatical sentences without the need for separate subjects—for example, *Tengo zapatos* ("I have shoes"), in which the person who has the shoes, "I," is indicated not by a separate word but by the "o" at the end of the verb. Chomsky contended that as soon as children encountered a few sentences of this type, their brains would set a switch to "on," indicating that the sentence subject should be dropped. Then they would know that they could drop the subject in all their sentences.

The "subject-drop" parameter supposedly also determined other structural features of the language. This notion of universal principles fits many European languages reasonably well. But data from non-European languages turned out not to fit the revised version of Chomsky's theory. Indeed, the research that had attempted to identify parameters, such as the subject-drop, ultimately led to the abandonment of the second incarnation of universal grammar because of its failure to stand up to scrutiny.

More recently, in a famous paper published in *Science* in 2002, Chomsky and his co-authors described a universal grammar that included only one feature, called computational recursion (although many advocates of universal grammar still prefer to assume there are many universal principles and parameters). This new shift permitted a limited number of words and rules to be combined to make an unlimited number of sentences.

The endless possibilities exist because of the way recursion embeds a phrase *within* another phrase of the same type. For example, English can embed phrases to the right ("John hopes Mary knows Peter is lying") or embed centrally ("The dog that the cat that the boy saw chased barked"). In theory, it is possible to go on embedding these phases infinitely. In practice, understanding starts to break down when the phrases are stacked on top of one another as in these examples. Chomsky thought this breakdown was not directly related to language per se. Rather it was a limitation of human memory. More important, Chomsky proposed that this recursive ability is what sets language apart from other types of thinking such as categorization and perceiving the relations among things. He also proposed recently this ability arose from a single genetic mutation that occurred between 100,000 and 50,000 years ago.

As before, when linguists actually went looking at the variation in languages across the world, they found counterexamples

© 2018 Scientific American

to the claim that this type of recursion was an essential property of language. Some languages—the Amazonian Pirahã, for instance—seem to get by without Chomskyan recursion.

As with all linguistic theories, Chomsky's universal grammar tries to perform a balancing act. The theory has to be simple enough to be worth having. That is, it must predict some things that are not in the theory itself (otherwise it is just a list of facts). But neither can the theory be so simple that it cannot explain things it should. Take Chomsky's idea that sentences in all the world's languages have a "subject." The problem is that the concept of a subject is more like a "family resemblance" of features than a neat category. About 30 different grammatical features define the characteristics of a subject. Any one language will have only a subset of these features—and the subsets often do not overlap with those of other languages.

Chomsky tried to define the components of the essential tool kit of language—the kinds of mental machinery that allow human language to happen. Where counterexamples have been found, some Chomsky defenders have responded that just because a language lacks a certain tool—recursion, for example—does not mean that it is not in the tool kit. In the same way, just because a culture lacks salt to season food does not mean salty is not in its basic taste repertoire. Unfortunately, this line of reasoning makes Chomsky's proposals difficult to test in practice, and in places they verge on the unfalsifiable.

The brain's innate sentence-diagramming machine, according to Chomsky, would fit words into correct grammatical slots—"nice" (adjective), "dogs" (noun).

DEATH KNELLS

A KEY FLAW in Chomsky's theories is that when applied to language learning, they stipulate that young children come equipped with the capacity to form sentences using abstract grammatical rules. (The precise ones depend on which version of the theory is invoked.) Yet much research now shows that language acquisition does not take place this way. Instead young children begin by learning simple grammatical patterns; then, gradually, they intuit the rules behind them bit by bit.

Thus, young children initially speak with only concrete and simple grammatical constructions based on specific patterns of words: "Where's the X?"; "I wanna X"; "More X"; "It's an X"; "I'm X-ing it"; "Put X here"; "Mommy's X-ing it"; "Let's X it"; "Throw X"; "X gone"; "Mommy X"; "I Xed it"; "Sit on the X"; "Open X"; "X here"; "There's an X"; "X broken." Later, children combine these early patterns into more complex ones, such as "Where's the X that Mommy Xed?"

Many proponents of universal grammar accept this characterization of children's early grammatical development. But then they assume that when more complex constructions emerge, this new stage reflects the maturing of a cognitive capacity that uses universal grammar and its abstract grammatical categories and principles.

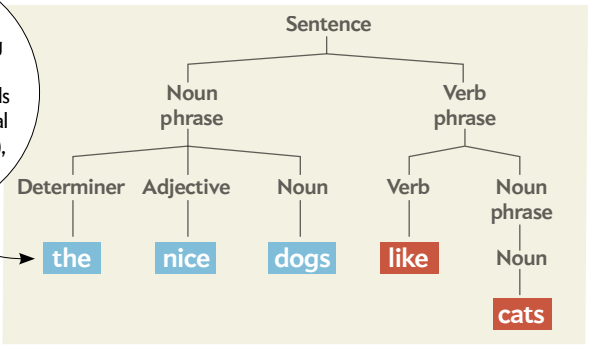
For example, most universal grammar approaches postulate that a child forms a question by following a set of rules based on grammatical categories such as "What (object) did (auxiliary) you (subject) lose (verb)?" Answer: "I (subject) lost (verb) something (object)." If this postulate is correct, then at a given developmental period children should make similar errors across all wh-question sentences alike. But children's errors do not fit this

Noam-enclature

Noam Chomsky took the linguistics community by storm more than 50 years ago. The idea was simple. Underlying language is a set of rules innate to every child that generates grammatical sentences from the earliest age. Chomsky set out to define those rules and how they work. Without this universal grammar, he thought, it would be impossible for a child to learn any language. In the ensuing years, Chomsky's theory has gradually been challenged by new theories asserting that language is acquired as children discern patterns in the language they hear around them.

Chomsky's Universal Grammar

Chomsky's universal grammar equipped the child with rules that worked on phrases ("the nice dogs") and rules for transforming those phrases ("Cats are liked by the nice dogs"). The theory has evolved in recent years but still retains the essential idea that children are born with the ability to make words conform to a grammatical template.



Usage-Based Learning

New approaches to linguistics and psychology suggest that children's natural ability to intuit what others think, combined with powerful learning mechanisms in the developing brain, diminishes the need for a universal grammar. Through listening, the child learns patterns of usage that can be applied to different sentences. The word "food" might replace the word "ball" after the phrase "The dog wants." Studies show that this theory of building up knowledge of word meaning and grammar approximates the way that two- and three-year-olds actually learn language.

Illustration by Lucy Reading-Ikkanda

© 2018 Scientific American

prediction. Many of them early in development make errors such as "Why he can't come?" but at the same time as they make this error—failing to put the "can't" before the "he"—they correctly form other questions with other wh-words and auxiliary verbs, such as the sentence "What does he want?"

Experimental studies confirm that children produce correct question sentences most often with particular wh-words and auxiliary verbs (often those with which they have most experience, such as "What does …"), while continuing to make errors with question sentences containing other (often less frequent) combinations of wh-words and auxiliary verbs: "Why he can't come?"

The main response of universal grammarians to such findings is that children have the competence with grammar but that other factors can impede their performance and thus both hide the true nature of their grammar and get in the way of studying the "pure" grammar posited by Chomsky's linguistics. Among the factors that mask the underlying grammar, they say, include immature memory, attention and social capacities.

Yet the Chomskyan interpretation of the children's behavior is not the only possibility. Memory, attention and social abilities may not mask the true status of grammar; rather they may well be integral to building a language in the first place. For example, a study co-authored by one of us (Ibbotson) showed that children's ability to produce a correct irregular past tense verb—such as "Every day I fly, yesterday I *flew*" (not "flyed")—was associated with their ability to inhibit a tempting response that was unrelated to grammar. (For example, to say the word "moon" while looking at a picture of the sun.) Rather than memory, mental analogies, attention and reasoning about social situations getting in the way of children expressing the pure grammar of Chomskyan linguistics, those mental faculties may explain why language develops as it does.

IN THE NEW USAGE-BASED APPROACH, CHILDREN ARE NOT BORN WITH A UNIVERSAL, DEDICATED TOOL FOR THE LEARNING OF GRAMMAR. INSTEAD THEY INHERIT THE MENTAL EQUIVALENT OF A SWISS ARMY KNIFE.

As with the retreat from the cross-linguistic data and the tool-kit argument, the idea of performance masking competence is also pretty much unfalsifiable. Retreats to this type of claim are common in declining scientific paradigms that lack a strong empirical base—consider, for instance, Freudian psychology and Marxist interpretations of history.

Even beyond these empirical challenges to universal grammar, psycholinguists who work with children have difficulty conceiving theoretically of a process in which children start with the same algebraic grammatical rules for all languages and then proceed to figure out how a particular language—whether English or Swahili—connects with that rule scheme. Linguists call this conundrum the linking problem, and a rare systematic attempt to solve it in the context of universal grammar was made by Harvard University psychologist Steven Pinker for sentence subjects. Pinker's account, however, turned out not to agree with data from child development studies or to be applicable to grammatical categories other than subjects. And so the linking problem—which should be the central problem in applying universal grammar to language learning—has never been solved or even seriously confronted.

AN ALTERNATIVE VIEW

ALL OF THIS LEADS ineluctably to the view that the notion of universal grammar is plain wrong. Of course, scientists never give up on their favorite theory, even in the face of contradictory evidence, until a reasonable alternative appears. Such an alternative, called usage-based linguistics, has now arrived. The theory, which takes a number of forms, proposes that grammatical structure is not innate. Instead grammar is the product of history (the processes that shape how languages are passed from one generation to the next) and human psychology (the set of social and cognitive capacities that allow generations to learn a language in the first place). More important, this theory proposes that language recruits brain systems that may not have evolved specifically for that purpose and so is a different idea to Chomsky's single-gene mutation for recursion.

In the new usage-based approach (which includes ideas from functional linguistics, cognitive linguistics and construction grammar), children are not born with a universal, dedicated tool for learning grammar. Instead they inherit the mental equivalent of a Swiss Army knife: a set of general-purpose tools—such as categorization, the reading of communicative intentions, and analogy making, with which children build grammatical categories and rules from the language they hear around them.

For instance, English-speaking children understand "The cat ate the rabbit," and by analogy they also understand "The goat tickled the fairy." They generalize from hearing one example to another. After enough examples of this kind, they might even be able to guess who did what to whom in the sentence "The gazzer mibbed the toma," even though some of the words are literally nonsensical. The grammar must be something they discern beyond the words themselves, given that the sentences share little in common at the word level.

The meaning in language emerges through an interaction between the potential meaning of the words themselves (such as the things that the word "ate" can mean) and the meaning of the grammatical construction into which they are plugged. For example, even though "sneeze" is in the dictionary as an intransitive verb that only goes with a single actor (the one who sneezes), if one forces it into a ditransitive construction—one able to take both a direct and indirect object—the result might be "She sneezed him the napkin," in which "sneeze" is construed as an action of transfer (that is to say, she made the napkin go to him). The sentence shows that grammatical structure can make as strong a contribution to the meaning of the utterance as do the words. Contrast this idea with that of Chomsky, who argued there are levels of grammar that are free of meaning entirely.

© 2018 Scientific American

The concept of the Swiss Army knife also explains language learning without any need to invoke two phenomena required by the universal grammar theory. One is a series of algebraic rules for combining symbols—a so-called core grammar hardwired in the brain. The second is a lexicon—a list of exceptions that cover all of the other idioms and idiosyncrasies of natural languages that must be learned. The problem with this dual-route approach is that some grammatical constructions are partially rule-based and also partially not—for example, "Him a presidential candidate?!" in which the subject "him" retains the form of a direct object but with the elements of the sentence not in the proper order. A native English speaker can generate an infinite variety of sentences using the same approach: "Her go to ballet?!" or "That guy a doctor?!" So the question becomes, are these utterances part of the core grammar or the list of exceptions? If they are not part of a core grammar, then they must be learned individually as separate items. But if children can learn these part-rule, part-exception utterances, then why can they not learn the rest of language the same way? In other words, why do they need universal grammar at all?

In fact, the idea of universal grammar contradicts evidence showing that children learn language through social interaction and gain practice using sentence constructions that have been created by linguistic communities over time. In some cases, we have good data on exactly how such learning happens. For example, relative clauses are quite common in the world's languages and often derive from a meshing of separate sentences. Thus, we might say, "My brother.... He lives over in Arkansas.... He likes to play piano." Because of various cognitive-processing mechanisms—with names such as schematization, habituation, decontextualization and automatization—these phrases evolve over long periods into a more complex construction: "My brother, who lives over in Arkansas, likes to play the piano." Or they might turn sentences such as "I pulled the door, and it shut" gradually into "I pulled the door shut."

What is more, we seem to have a species-specific ability to decode others' communicative intentions—what a speaker intends to say. For example, I could say, "She gave/bequeathed/sent/loaned/sold the library some books" but not "She donated the library some books." Research has now shown that there are several mechanisms that lead children to constrain these types of inappropriate analogies. For example, children do not make analogies that make no sense. So they would never be tempted to say "She ate the library some books." In addition, if children hear quite often "She donated some books to the library," then this usage preempts the temptation to say "She donated the library some books."

Such constraining mechanisms vastly cut down the possible analogies a child could make to those that align the communicative intentions of the person he or she is trying to understand. We all use this kind of intention reading when we understand "Can you open the door for me?" as a request for help rather than an inquiry into door-opening abilities.

Chomsky allowed for this kind of "pragmatics"—how we use language in context—in his general theory of how language worked. Given how ambiguous language is, he had to. But he appeared to treat the role of pragmatics as peripheral to the main job of grammar. In a way, the contributions from usage-based approaches have shifted the debate in the other direction to how much pragmatics can do for language before speakers need to turn to the rules of syntax.

Usage-based theories are far from offering a complete account of how language works. Meaningful generalizations that children make from hearing spoken sentences and phrases are not the whole story of how children construct sentences either—there are generalizations that make sense but are not grammatical (for example, "He disappeared the rabbit"). Out of all the possible meaningful yet ungrammatical generalizations children could make, they appear to make very few. The reason seems to be they are sensitive to the fact that the language community to which they belong conforms to a norm and communicates an idea in just "this way." They strike a delicate balance, though, as the language of children is both creative ("I goed to the shops") and conformative to grammatical norms ("I went to the shops"). There is much work to be done by usage-based theorists to explain how these forces interact in childhood in a way that exactly explains the path of language development.

A LOOK AHEAD

AT THE TIME the Chomskyan paradigm was proposed, it was a radical break from the more informal approaches prevalent at the time, and it drew attention to all the cognitive complexities involved in becoming competent at speaking and understanding language. But at the same time that theories such as Chomsky's allowed us to see new things, they also blinded us to other aspects of language. In linguistics and allied fields, many researchers are becoming ever more dissatisfied with a totally formal language approach such as universal grammar—not to mention the empirical inadequacies of the theory. Moreover, many modern researchers are also unhappy with armchair theoretical analyses, when there are large corpora of linguistic data—many now available online—that can be analyzed to test a theory.

The paradigm shift is certainly not complete, but to many it seems that a breath of fresh air has entered the field of linguistics. There are exciting new discoveries to be made by investigating the details of the world's different languages, how they are similar to and different from one another, how they change historically, and how young children acquire competence in one or more of them.

Universal grammar appears to have reached a final impasse. In its place, research on usage-based linguistics can provide a path forward for empirical studies of learning, use and historical development of the world's 6,000 languages. SA

Paul Ibbotson is a lecturer in language development at the Open University, based in England.

Michael Tomasello is co-director of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, and author, most recently, of *A Natural History of Human Morality* (Harvard University Press, 2016).

MORE TO EXPLORE

Constructing a Language: A Usage-Based Theory of Language Acquisition. Michael Tomasello. Harvard University Press, 2003.

Constructions at Work: The Nature of Generalization in Language. Adele Goldberg. Oxford University Press, 2006.

Language, Usage and Cognition. Joan Bybee. Cambridge University Press, 2010.

scientificamerican.com/magazine/sa

© 2018 Scientific American

The Puzzle of Dark Energy

Why is the expansion of the universe accelerating? After two decades of study, the answer is as mysterious as ever, but the questions have become clearer

By Adam G. Riess and Mario Livio

THE UNIVERSE IS GETTING BIGGER EVERY SECOND. GALAXIES ARE FLYING apart from one another, clusters of galaxies are zooming away from other clusters, and the empty space between everything is growing wider and wider. This much was known since the 1920s, when observations by Edwin Hubble and others revealed that the cosmos is expanding. But more recently, astronomers found that the process is speeding up—the pace of the expansion of the universe is rising, so that galaxies are receding from one another faster now than they were a moment ago. This is the startling realization that one of us (Riess), along with collaborators he co-led with Brian Schmidt of the Australian National University, came to in 1998 through measurements of distant supernovae explosions.

Illustration by Kenn Brown, Mondolithic Studios

© 2018 Scientific American

LIKE A RUBIK'S CUBE, the quandary of dark energy—the mystery of why the expansion of the universe is speeding up—is turning out to be hard to crack.

© 2018 Scientific American

The discovery agreed with findings from another team led by Saul Perlmutter of the University of California, Berkeley, using a similar method published the same year. The conclusion was unavoidable—something was causing the expansion of the universe to pick up speed. But what?

We give the name "dark energy" to whatever is causing the repulsive force that is apparently pulling the universe apart. After studying the situation for two decades, the physical nature of dark energy remains almost as elusive today as it was 20 years ago.

We are left with several pressing mysteries: What is dark energy? Why does it seem so much weaker than the most straightforward theories predict (and yet strong enough that we detect it)? What does the nature of dark energy mean for the future of the universe? And finally, do the strange characteristics of dark energy imply that our universe got its properties randomly—that in fact, our universe is just one of a vast multiverse containing untold versions of the cosmos, each with different traits and different strengths of dark energy?

An all-out assault to identify the nature of dark energy is under way, and prospects are bright for several new observatory projects to make progress soon. Within the next decade we hope to begin to answer these questions and comprehend the nature of cosmic acceleration—or resign ourselves to leaving some mysteries unsolved indefinitely.

WHAT IS DARK ENERGY?

SCIENTISTS HAVE A NUMBER of hypotheses for what might be driving the acceleration of the universe. The leading candidate arises from the nature of empty space. In quantum physics a vacuum is not "nothing"—rather it is teeming with pairs of "virtual" particles and antiparticles that spontaneously appear and annihilate one another within a tiny fraction of a second. As strange as it may sound, this sea of ephemeral particle pairs carries energy, and energy, just like mass, can produce gravity. Unlike mass, however, energy can create either an attractive or a repulsive gravity, depending on whether its pressure is positive or negative. The vacuum energy in empty space, according to theory, should have a negative pressure and thus may be the source of the repulsive gravity driving the accelerated expansion of the universe.

This idea is equivalent to the "cosmological constant," a term Albert Einstein added to his general relativity equations that represents a constant energy density throughout space. As the name implies, this hypothesis holds that the density of dark energy is constant—that is, unvarying—over space and time. So far the astrophysical evidence we have best fits with the cosmological constant explanation, with some discrepancies.

Alternatively, dark energy may be an energy field dubbed "quintessence" that pervades the universe, imbuing every point in space with a property that counteracts the pull of gravity. Physicists are familiar with fields—the everyday forces of electromagnetism and gravity act via fields (although these usually arise from localized sources and do not pervade all of space).

If dark energy is a field, it would not be a constant and so might change over time. In that case, dark energy might once have been stronger or weaker than it is now and could have affected the universe differently at different times. Likewise, its strength and impact on the evolution of the universe might alter in the future. In the so-called freezing-field version of this idea, dark energy evolves more and more slowly as time progresses; in the thawing variant, the field changes slowly at first and faster later.

A third option may account for cosmic acceleration: there is no dark energy, and the quickening expansion of the universe results from physics not explained by Einstein's theory of gravity (general relativity), which is incomplete. It is possible that in truly extreme regimes, such as the breadth of galaxy clusters or the entire observable universe, the laws of gravity work differently than the theory predicts, and gravity misbehaves. Physicists have put forth a few interesting theoretical suggestions along these lines, but no self-consistent theory that agrees with all the observations currently exists, so dark energy seems to have the upper hand over this option for now. (Previous ideas, such as the notion that cosmic acceleration is a manifestation of an uneven distribution of matter throughout the universe or the result of a network of geometric defects in the structure of space, have by now largely proved to be inconsistent with observational data.)

WHY IS DARK ENERGY SO WEAK?

NONE OF THE PROPOSED explanations for dark energy is very satisfying. The cosmological constant, for example, predicts that dark energy should be vastly stronger than it actually is. When one naively attempts to sum up the energies over all the presumed quantum states associated with the sea of virtual particles and antiparticles in the vacuum of space, one obtains a value that is more than 120 orders of magnitude larger than the observed sum. Factoring in ideas from proposed theories such as supersymmetry—the notion that every known particle has a heavier partner particle that we have not yet discovered—reduces the discrepancy somewhat, but the difference between the predicted and the measured total energy still remains tens of orders of magnitude too high. So if dark energy is explained by the energy of the vacuum, the question is, How did this vacuum energy turn out to be so tiny?

The field explanation for dark energy barely does better on this front. Theorists simply assume (without a good explanation for why it should be so) that the minimum of the potential energy associated with the dark energy field is very low, guaranteeing that only a small amount of dark energy is spread throughout space. Also, these models require the field to have surprisingly minimal interactions with everything else in the universe (apart from its repulsive gravitational push)—a property that is hard to explain. These facts make it difficult to incorporate the

IN BRIEF

Twenty years ago scientists discovered that the expansion of the universe is accelerating and dubbed the source of this acceleration "dark energy."
Intensive investigations since then have not resolved the nature of dark energy but rather exposed a number of further questions: Why is dark energy so much weaker than theory seems to predict, what does it mean for the future of the cosmos, and might dark energy lead to the conclusion that we live in a multiverse?
With the advent of several recent and upcoming experiments, scientists hope that the coming years will finally bring some answers.

© 2018 Scientific American

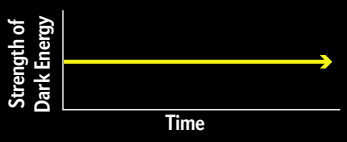
Dark Energy Possibilities and Potential Futures

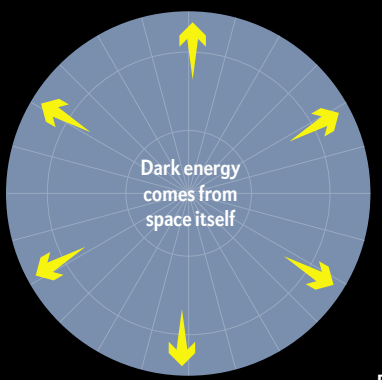
Dark energy is scientists' name for whatever is causing the expansion of the universe to accelerate. Explanations for dark energy fall into three main categories: it may be an unchanging energy arising from empty space (an idea called the cosmological constant) or a varying energy stemming from a field pervading the universe (quintessence). Or dark energy may not exist at all—in that case, gravity would act differently than thought on cosmic scales.

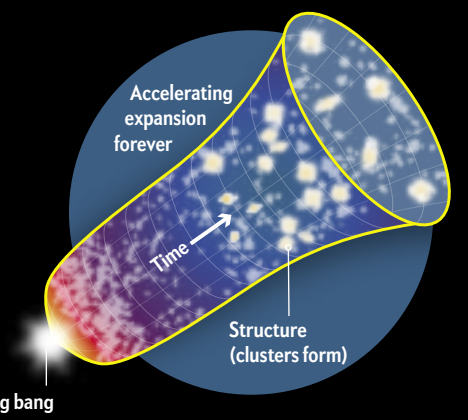
MODEL → FUTURE

Cosmological Constant

If the vacuum of empty space has an inherent energy, it may push the universe to expand. The strength of such an energy would be constant through time and would act just like the cosmological constant term Albert Einstein added to, and later removed from, his equations of general relativity.

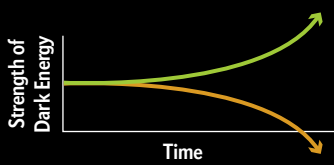


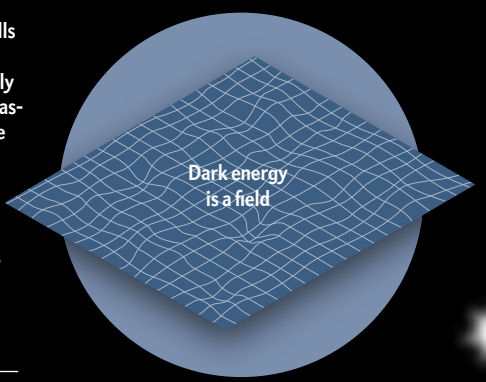


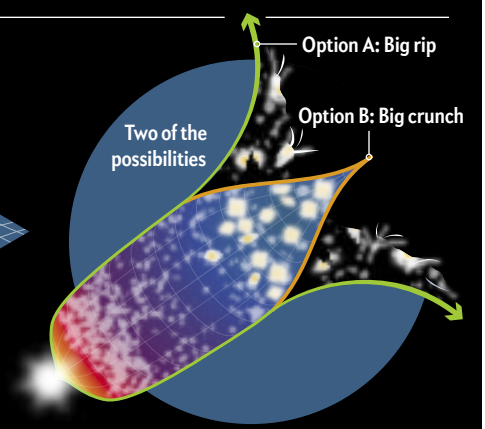


Quintessence

If dark energy comes from a field that fills the cosmos, its strength could change over time, either increasing to eventually rip all structures in space apart or decreasing and changing directions to allow the universe to contract in a big crunch.





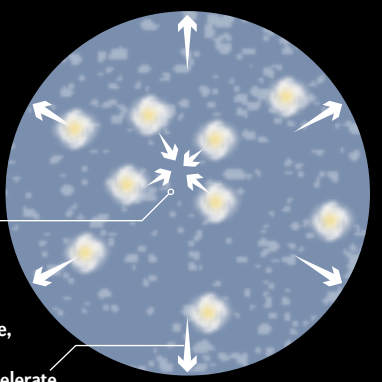


There Is No Dark Energy

Dark energy may not exist at all, and the acceleration of the universe's expansion may instead indicate that gravity operates differently than we think on extremely large scales.

On the scale of galaxies and clusters, gravity behaves as general relativity predicts

On the scale of the universe as a whole, gravity grossly diverges from general relativity; the universe appears to accelerate



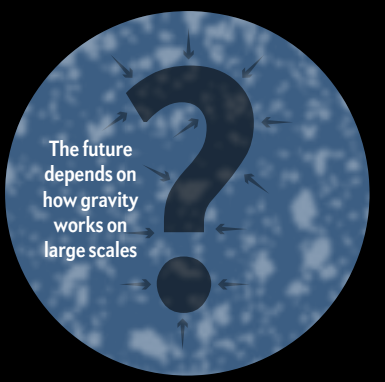
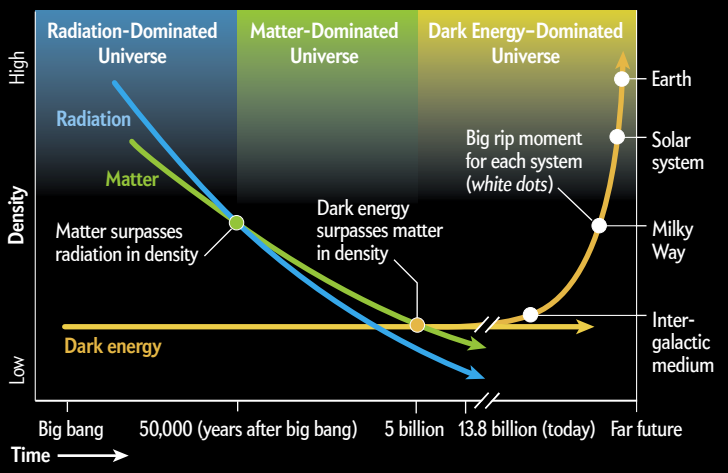


Illustration by Nigel Hawtin

© 2018 Scientific American

Cosmic Control

Because dark energy is denser in space than any other constituent of the universe, it exerts the dominant influence on the cosmos and will therefore control its fate. Dark energy was not always on top, though: the other ingredients of the universe—radiation (light) and matter (including atoms and regular matter as well as invisible dark matter)—were dominant when the universe was young and small, and they were packed tightly in space. As the universe expanded over time, matter and radiation spread out, and dark energy overpowered them. If the density of dark energy increases, it may become so powerful that it rips apart all structures in space.



dark energy field hypotheses naturally within common particle physics models.

WHAT DOES THIS MEAN FOR THE FUTURE OF THE UNIVERSE?

THE PROPERTIES of dark energy will determine the ultimate fate of our universe. For instance, if dark energy is indeed the energy of empty space (the cosmological constant), the acceleration will continue forever, and about a trillion years from now expansion will cause all the galaxies that are more distant than the Milky Way's closest neighbors (the Local Group, which by then will coalesce to form one large elliptical galaxy) to separate faster than light speed, rendering them undetectable. Even the ancient light from the afterglow of the big bang—the cosmic microwave background (CMB) that fills all of space—will be stretched to wavelengths longer than the size of the visible universe and thus rendered imperceptible. In this scenario, we happen to be living at a very fortunate time when we still have the best possible view of our universe.

If, on the other hand, dark energy is not the energy of the vacuum but rather the energy of some unknown field, then the future is wide open. Depending on the way the field evolves, the universe could eventually stop expanding and start collapsing, falling in on itself to a final "big crunch" that mimics the big bang from whence it came. Or the universe could end up in a "big rip," in which all complex structures, from clusters of galaxies to atoms and atomic nuclei, could become overwhelmed by dark energy and tear apart. And the first scenario, continuous acceleration toward a cold death, is also an option with a dark energy field.

An alternative theory of gravity, if it turns out to be necessary, likewise allows for various outcomes depending on the particularities of the revised theory.

MIGHT WE LIVE IN A MULTIVERSE?

WITH THE COSMOLOGICAL CONSTANT explanation leading the theoretical pack, the problem of its inexplicable weakness comes to the forefront. Realizing this problem with the constant even before the discovery of accelerating expansion, physicist Steven Weinberg of the University of Texas at Austin suggested a new paradigm—one in which the cosmological constant is not uniquely determined from the basic laws of physics but rather is a random variable that assumes different values in different members of a huge ensemble of universes—a multiverse. Some universes may have much larger cosmological constants, but in those the accelerating repulsive force is so large that matter cannot coalesce to form galaxies, planets and life. Because we exist, Weinberg reasons, we clearly have to find ourselves in one of those universes that *can* allow for our existence—one that happened to have a small cosmological constant. This idea, which was further developed by Alexander Vilenkin of Tufts University, Martin Rees of the University of Cambridge, one of us (Livio) and others, is called anthropic reasoning.

There are good reasons, aside from the dark energy consideration, that a multiverse might arise. The widely accepted theory of cosmic inflation suggests the universe ballooned stupendously in its first fraction of a second. Vilenkin and Andrei Linde of Stanford University have shown that once cosmic inflation starts, it is essentially impossible to stop it from occurring again and again, thereby creating an infinite ensemble of bubbles or "pocket universes" that form in isolation from one another and may have very different properties.

A multiverse also seems to be a consequence of string theory, a candidate for a theory that unifies all the forces of nature. Calculations based on versions of string theory called M-theory by Raphael Bousso and the late Joseph Polchinski suggest that there could be as many as 10^{500} different spacetimes, or universes, each characterized by different values for the constants of nature and even the number of dimensions in space.

Even just mentioning the multiverse idea, however, raises the blood pressure of some physicists. The notion seems hard to swallow and harder to test—perhaps signifying the end of the classical scientific method as we know it. Historically this method has required that hypotheses should be directly testable by new experiments or observations. Yet the concept of the multiverse does make a few predictions that may yield to testing. In particular, some multiverse models predict that the shape of space-

Graphic by Jen Christiansen

© 2018 Scientific American

time has a slight curve to it that might be detectable by observations. Another possibility, albeit not a very likely one, is that the cosmic microwave background light may contain ripples that are the signature of a collision of another bubble with our own.

FINDING ANSWERS

THE BEST WAY we know of to start to reveal the nature of dark energy is to measure the ratio of its pressure (how much it is in tension with space) to its density (how much of it is in a given amount of space)—a property called its equation of state parameter, w. If dark energy is the energy of the vacuum (the cosmological constant), then w will be constant and equal to –1. If dark energy is associated with a field that changes with time, on the other hand, we would hope to detect its w value differing from –1 and evolving throughout cosmic history. Alternatively, if the observed acceleration necessitates a modification of Einstein's theory of gravity for extreme distances, we would hope to observe an inconsistency between the value of w we find at different scales in the universe.

Astronomers have devised some clever indirect ways to measure dark energy's pressure and density. As a repulsive gravitational pull, dark energy or modified gravity would counteract the pull of regular gravity (which tugs mass in the universe toward other mass), discouraging the formation of large-scale structure—that is, galaxy clusters. Thus, by studying how clusters grew over time, scientists can find out how strong dark energy was at various points in history. We do this by observing how the mass of clusters bends the light of background galaxies behind them through a process known as gravitational lensing. The amount of bending tells us how massive the clusters are, and by observing this effect for clusters at varying distances, we can measure how common massive clusters were at various cosmic epochs. (Because of the time it takes light to reach us, looking far away is akin to looking back in time.)

We can also measure dark energy by studying how the rate of the universe's expansion has changed over time. By observing objects at different distances and measuring their redshift—how much the wavelength of their light has been stretched by the expansion of space—we can learn how much the universe has expanded since that light began its journey. This method, in fact, was how the two teams initially discovered cosmic acceleration; they measured the redshifts of different type 1a supernovae (whose distances are reliably tied to their brightness). A variation on this technique is to observe the apparent size of ripples in the density of galaxies across space called baryon acoustic oscillations (BAOs)—another reliable distance indicator—as a way to trace the expansion history of the universe.

To date, most measurements of w are generally consistent (within the observational uncertainties) with a value of –1 to within 10 percent and thus support the cosmological constant explanation of accelerated expansion. When a team led by Riess used the Hubble Space Telescope to probe dark energy back to about 10 billion years ago using the supernova technique, it found no evidence for variation over that time. Similarly, an analysis of the results from a sample of more than 1,000 supernovae from several surveys also gave results that are consistent with a cosmological constant.

It is worth noting, however, that some hints of a possible deviation from cosmological constant predictions have shown up in the past couple of years. For instance, a combination of measurements of the CMB (which tells us about the total mass and energy in the universe) from the Planck satellite with results from gravitational-lensing studies suggests a value for w that is more negative than –1. Very recent observations of baryon acoustic oscillations in data from distant bright galaxies called quasars seem to hint that dark energy density has increased with time. Finally, a small (but statistically significant) discrepancy between local measurements of the rate of space's expansion today, compared with measurements of the primordial rate of expansion from the CMB, could also be pointing to a deviation from a cosmological constant. As intriguing as these results are, none is yet compelling evidence of a departure from a cosmological constant. More data in the near future may strengthen these discrepancies and indicate some new physics or reveal them to be systematic flukes.

Work is now under way to achieve a 100-fold improvement in the precision of the measured properties of dark energy in the coming decade. New projects such as the Dark Energy Survey (DES) begun in 2013 and the Large Synoptic Survey Telescope (LSST) due to open around 2021 will gather better information about large-scale structure in the universe and the history of expansion. NASA's Wide-Field Infrared Survey Telescope–Astrophysics Focused Telescope Assets (WFIRST-AFTA), a 2.4-meter space telescope planned to launch in the mid-2020s, is expected to observe distant supernovae and BAOs, as well as gravitational lensing. The launch of the European Space Agency's Euclid space mission is currently planned for 2021 and will also exploit lensing, BAOs and redshift measurements of galactic distances to determine the three-dimensional distribution of galaxy clusters.

Finally, we can also test theories of modified gravity through measurements within the solar system. One method measures the distance to the moon to such an astonishingly high precision (through reflecting laser light off reflectors placed on the moon by *Apollo* astronauts) that it can detect minute deviations from predictions of general relativity. In addition, ingenious laboratory experiments will search for minuscule discrepancies in the current laws of gravity.

The coming years should be a pivotal time for dark energy research. We have hopes of being able to make real progress in addressing the outstanding questions about the accelerating expansion of the universe. The answers will reveal nothing short of the future of the cosmos. SA

Adam G. Riess is an astrophysicist at Johns Hopkins University and the Space Telescope Science Institute. His research on distant supernovae revealed that the expansion of the universe is accelerating, a discovery for which he shared the 2011 Nobel Prize in Physics.

Mario Livio is an astrophysicist who worked for 24 years with the Hubble Space Telescope and a best-selling author of popular science books; his most recent one is *Why?: An Exploration of Human Curiosity* (Simon & Schuster, 2017).

MORE TO EXPLORE

Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. Adam G. Riess et al. in *Astronomical Journal,* Vol. 116, No. 3, pages 1009–1038; September 1998.

The Accelerating Universe. Mario Livio. Wiley, 2000.

scientificamerican.com/magazine/sa

© 2018 Scientific American

THE RED LINE

Scientists are on the threshold of crossing a sharp ethical line: permanently altering the human genetic code

By Stephen S. Hall

Kyle Orwig has been itching to do an experiment that would, in his words, "piss people off." Orwig, a professor at the University of Pittsburgh, is an expert on the intricate biology of sperm cells—in particular, how specialized "stem" cells located in the male testes produce sperm. Every so often, however, a genetic flaw prevents these stem cells from completing this process, thus rendering the male infertile. The experiment Orwig has in mind is to use gene-editing technology to fix this flaw in the sperm-forming stem cells and then transplant them back into infertile mice, thereby demonstrating a potential treatment for male infertility.

It sounds simple enough, and, according to Orwig, it would be relatively straightforward to try—indeed, he has been transplanting sperm-forming stem cells into mice for 20 years. The consequences, however, could be momentous. The kind of experiment Orwig is contemplating would, if successful, push society right up to the brightest red line in contemporary biology: altering the genetic text of the human species in a way that is passed down to future generations.

If shown to be safe, effective and ethically acceptable, germline modification would confer unprecedented power on scientists—the power to edit the susceptibility to disease out of our species' DNA, for example, but also the power to manipulate human inheritance and "improve" the species, an aim that darkly harkens back to the eugenics movement of the early 20th century that reached its nadir in Nazi Germany.

Orwig, a broad-shouldered, burr-headed Oregonian who seems genial and determined, has no intention of crossing ethical lines. But he is something of a provocateur. By demonstrating that infertility in mice could be cured with a modest amount of genetic tinkering, he hopes to trigger a wider awareness that editing human genes is not an abstract, long-term technical challenge, as some have suggested, but a near-term possibility with practical medical consequences. Which is why Orwig recently told a colleague, "Let's just do this and piss some people off. Show them that it's possible, so nobody can say it's impossible. And get people talking about it."

The issue of germ-line modification has assumed great urgency in the past two years because of a powerful gene-editing tool called CRISPR/Cas9, which allows scientists to alter the DNA of any organism—including, potentially, humans—with unprecedented precision and ease. In April 2015 Chinese researchers reported the first attempt to edit the genes of human embryos. The headlines—"Embryo Editing Sparks Epic Debate," in *Nature;* "Eugenics Lurk in the Shadow of CRISPR," in *Science*—signaled broad social unease. In the alarmist shorthand of media accounts, the possibility of gene editing triggered fears of "designer babies" and "genetic enhancement."

The humble sperm cell, however, is a less controversial target. And whereas editing genes in an embryo remains a big challenge, many experts believe that approaching germ-line modification upstream of the embryo, in the sex cells that merge to form a zygote, is easier and arguably safer. Once you modify these cells, however, you are essentially modifying the human genome because the changes are permanently inscribed in the genetic text of the embryos they create. Orwig is among a handful of biologists who have experience genetically modifying and transplanting spermatogonial stem cells, the testicular cells that churn out sperm.

IN BRIEF

New gene-editing techniques may soon be used in treatments for male infertility that involve altering the genetic code of sperm cells. Such alterations would be passed down to future generations, constituting a permanent change in the human genome. **Scientists seeking** to quell the public's fears that such a step would cross an ethical red line have argued that the technology is not yet feasible, but others believe the line will be crossed soon. **The demand** for infertility treatments has for decades fueled the adoption of new biotechnology in the clinic, for better and for worse.

Photograph by Andrew Myers

© 2018 Scientific American

© 2018 Scientific American

The field of reproductive medicine has a well-established track record for pushing headlong into the clinic with technological innovations. Infertility is also big business. If Orwig were to demonstrate in animals that a simple genetic fix is possible, it would be a tempting procedure for the tens of thousands of men who cannot make their own sperm, for whom options are currently limited, as well as for the in vitro fertilization (IVF) industry, which did an estimated $2 billion in business in 2015 in the U.S. (and perhaps 10 times that amount worldwide).

Before administering any treatment, there must be proof that it works and causes no undue harm. Scientists would require such proof before even contemplating creating a human being with edited genes. Animal versions of those experiments are already under way, however, and the red line may be crossed soon. It could happen in China, where researchers have already taken the first tentative steps in editing viable human embryos. It could happen in the U.K., where the government has legalized a form of germ-line modification known as mitochondrial replacement therapy and, in December 2016, approved the technique in fertility treatments using DNA from three people to prevent diseases from being passed on to children. And it could happen in an IVF clinic anywhere, building on recipes developed in laboratories such as Orwig's.

"This isn't theoretical," Orwig says. "The mouse is here, and the human is not too far in the future. The pieces are already in place."

KEEP CALM AND CARRY ON

THE CURRENT DEBATE over germ-line modification may feel familiar, but it treads on fundamentally new ground. Scientists began to acquire the Promethean ability to rewrite the language of heredity in the early 1970s, when biologists discovered that they could crudely cut and paste DNA with the use of enzymes harvested from bacteria, a technique called recombinant DNA. That advance caused unease about dangerous, genetically engineered microbes escaping the lab, prompting a voluntary—and unprecedented—moratorium on recombinant DNA research in 1974 and a historic meeting in 1975 of scientists at the Asilomar conference in California. Prominent molecular biologists such as David Baltimore, then at the Massachusetts Institute of Technology, debated the safety of the new technology, which led to federal guidelines governing the research. The Asilomar meeting was correctly viewed as a watershed cultural moment: Michael Rogers published a richly detailed account of "The Pandora's Box Congress" in *Rolling Stone,* and by the time the guidelines were in place, biotechnology emerged as one of the transformative industries of the 20th century.

Although society applauded the scientific community's decision in 1974 to hit the pause button on the headlong rush of research, many scientists felt it was an overreaction to hypothetical safety concerns. James D. Watson, co-discoverer of the DNA double-helix structure, called it "senseless hysteria."

Since Asilomar, controversial biology has often landed in the public square with a bang, prompting big meetings with messy background noise. When the National Academy of Sciences debated recombinant DNA in 1977, protesters opposed to genetic engineering unfurled a banner quoting Adolf Hitler: "We Will Create the Perfect Race." A meeting on human cloning in 2001 became a media circus. Maverick IVF doctors vowed to clone human babies. Television crews followed the would-be cloners everywhere (including to the bathroom). *Wired* magazine, in 2001, proclaimed on its cover: "Someone Will Clone a Human in the Next 12 Months."

This time around there is palpable unease among scientists, but they are also wary that another self-imposed moratorium might hold up progress. The result? Another meeting. In December 2015 the National Academy of Sciences and the National Academy of Medicine co-hosted an international "summit" in Washington, D.C. (with the Royal Society and the Chinese Academy of Sciences). Baltimore acknowledged that altering human inheritance had remained "unthinkable" because of the cumbersome and inefficient first-generation tools of genetic engineering. "Over the years, however, the unthinkable has become conceivable," he said, "and today we sense that we are close to being able to alter human heredity." The overriding question, Baltimore continued, is, "How, if at all, do we as a society want to use this capability?"

The answer, to anyone who sat through the three-day meeting, as I did, seemed to be: we're not sure, but there's plenty of time to think it through. Numerous talks, including a plenary address by genome scientist Eric Lander of the Broad Institute of M.I.T. and Harvard University, stressed the technical hurdles and the lack of compelling medical needs for human germ-line modification in the near future. "It might, might, might be a good idea that before we make permanent changes to the human gene pool," Lander warned, "we should exercise considerable caution."

The organizers deftly steered clear of an Asilomar-like moratorium. Baltimore read a carefully worded statement from the meeting organizers acknowledging that it would be "irresponsible" to pursue human germ-line editing in the clinic at this time. At the close of the summit, he went on to explain that the organizers had deliberately avoided calling for a ban or moratorium. "Neither word did we want to use," he said. "Neither word have we used." Basic research could, and should, proceed unimpeded, but the public need not be concerned about imminent developments: human applications of germ-line editing were implausible, unnecessary, unwise and certainly not right around the corner.

That is not the way everyone in the scientific community sees it. The organizers of the Washington meeting framed the issue as "when, if at all." But a different word often crops up in private conversations with biologists when you ask about the prospects for germ-line editing. That word is "inevitable."

THE TIME LINE

SOME SCIENTISTS viewed the National Academies meeting as an effort to "reinforce the status quo," according to biologist George Church of Harvard Medical School. "They basically want the public to calm down," he says. "That was their goal. And no matter what we said, that was going to be the goal. I don't want to stir [the public] up, and I don't want to calm them down, either. I want them to have an accurate view of where things are going." And the public needs to start thinking about gene editing the human genome now, Church says, because science is already bumping up against the red line.

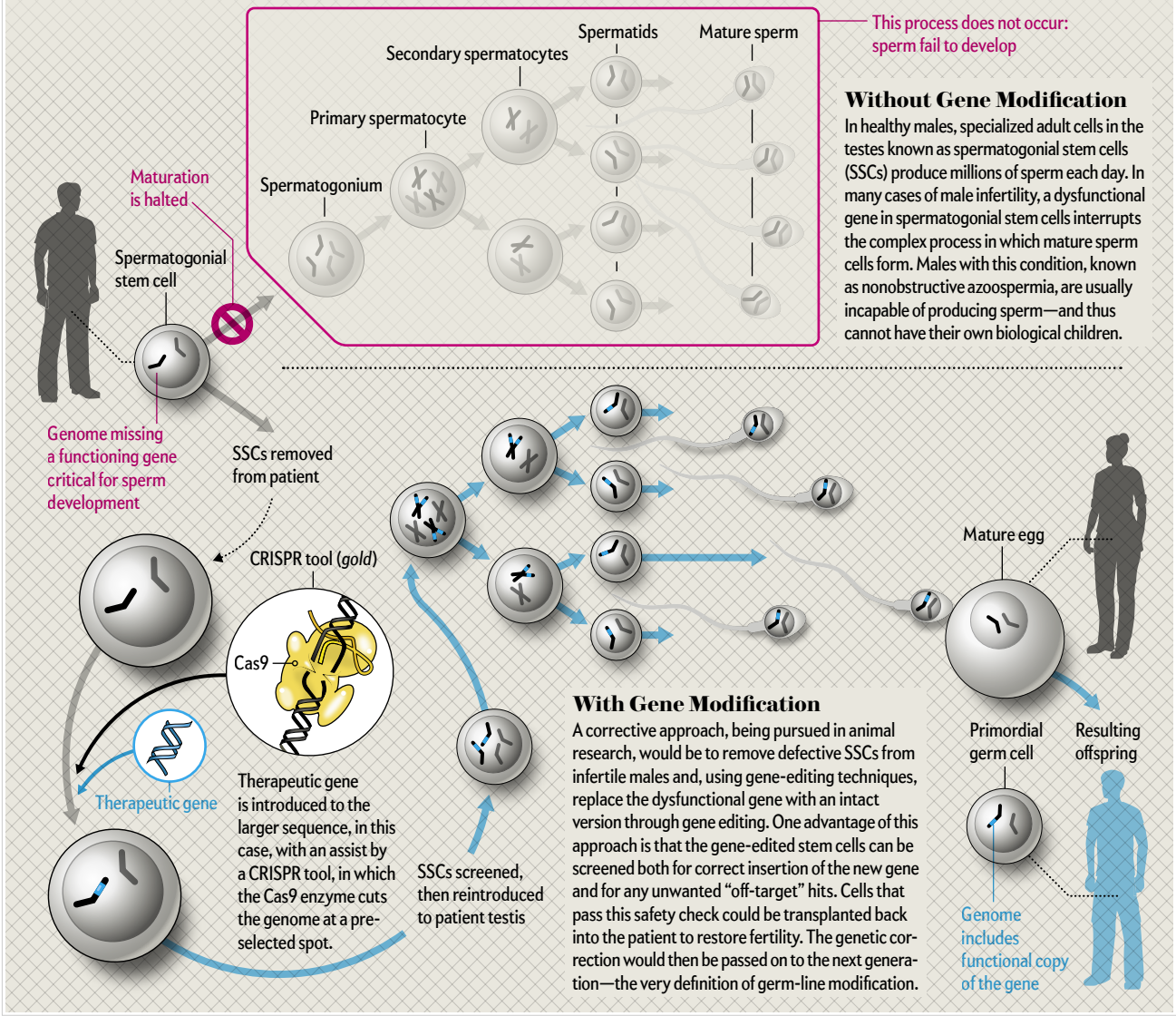
Despite a thicket of international provisions regulating human embryo research, Church and others believe that the creation of gene-edited sex cells in the test tube (the technical term is "in vitro gametogenesis," or IVG) has made great strides in recent years without attracting the same public scrutiny—or provoking the same ethical discomfort—as gene editing in embryos.

PRECEDING PAGE: PROP STYLING BY ANGELA CAMPOS *Stockland Martel*

© 2018 Scientific American

How to Modify the Germ Line

Scientists in China have reportedly tried to edit the genes of human embryos for research purposes. They are also exploring the possibility of modifying genes in sex cells (sperm or egg), which would produce changes that are passed to the next generation. One potential route is to modify genes in the stem cells that produce sperm, which could lead to treatments for both male infertility and single-gene disorders. (In 2015, for example, Chinese researchers used such an approach to fix a mutation that causes cataracts in mice.)


Spermatids
Mature sperm
Secondary spermatocytes
Primary spermatocyte
Spermatogonium
This process does not occur: sperm fail to develop
Maturation is halted
Spermatogonial stem cell
Genome missing a functioning gene critical for sperm development
SSCs removed from patient
CRISPR tool (gold)
Cas9
Therapeutic gene
Therapeutic gene is introduced to the larger sequence, in this case, with an assist by a CRISPR tool, in which the Cas9 enzyme cuts the genome at a pre-selected spot.
SSCs screened, then reintroduced to patient testis
Mature egg
Primordial germ cell
Resulting offspring
Genome includes functional copy of the gene

Without Gene Modification

In healthy males, specialized adult cells in the testes known as spermatogonial stem cells (SSCs) produce millions of sperm each day. In many cases of male infertility, a dysfunctional gene in spermatogonial stem cells interrupts the complex process in which mature sperm cells form. Males with this condition, known as nonobstructive azoospermia, are usually incapable of producing sperm—and thus cannot have their own biological children.

With Gene Modification

A corrective approach, being pursued in animal research, would be to remove defective SSCs from infertile males and, using gene-editing techniques, replace the dysfunctional gene with an intact version through gene editing. One advantage of this approach is that the gene-edited stem cells can be screened both for correct insertion of the new gene and for any unwanted "off-target" hits. Cells that pass this safety check could be transplanted back into the patient to restore fertility. The genetic correction would then be passed on to the next generation—the very definition of germ-line modification.

"In terms of the technology, that's ready to go now," says I. Glenn Cohen, a bioethicist at Harvard Law School. "In vitro gametogenesis is much closer to the mark than any other way to do it." And Ina Dobrinski, an expert in reproductive biology at the University of Calgary who works on gene editing of large animals such as pigs, adds, "Theoretically, we can do it. Practically [speaking], nobody is even touching it because of the ethical issues."

If human germ-line editing is inevitable, despite the ethical concerns (and indeed legal prohibitions in many countries), how might it happen? Speculation among biologists has become something of a parlor game, but I turned to Church, a card-carrying futurist, to lay out a plausible scenario. He was happy to oblige.

Church sees the germ-line Rubicon being crossed because sperm do not seem to arouse the same ethical passions as embryos or even egg cells. (Bioethicist Cohen agrees: "People do not believe that masturbation is genocide.") He also thinks gene therapy, not CRISPR per se, will set the table for this momentous change because it is already accepted: the U.S. Food and Drug Administration has allowed many trials of gene therapy in somatic (nongerm-line) cells. "Gene therapy is happening in young children already and will happen in younger and younger children," Church says.

Illustration by Jen Christiansen

© 2018 Scientific American

In a highly publicized case in fall 2015, for example, British researchers used gene-editing techniques to alter the immune cells of an infant battling leukemia. And the leap to germ-line gene therapy, according to Church, will occur not in human embryos but in the most proletarian, plentiful and expendable cell in the human body: sperm. Gene-editing sperm will spare couples the agony of destroying IVF embryos that, on preimplantation screening, are shown to possess variants that forecast some devastating single-gene disorders, he believes. "Maybe half the people in the U.S. already feel that they're not comfortable killing embryos, but I think people would be comfortable genetically altering sperm," he observes.

Two obvious targets, Church adds, would be single-gene disorders (such as Tay-Sachs disease) and infertility. "You could also do it in human spermatogonial stem cells," he says, referring to the specialized adult stem cells in male testes that generate sperm cells—millions on millions of mindless, headlong swimmers—every day. "People don't really care about spermatogonia. Most people don't even know how to pronounce it. So they're going to let you mess around with them, right?" Church continues. "You will be able to do all kinds of things to show that they are functionally normal—that you've taken the sperm that can't swim, and that you've now taken their stem cells and made sperm that can swim. You can test that in a lab without any eggs being involved. And then in the fertility clinic, the dad will say, 'Hey, those are pretty awesome sperm. Let's try them out and see what they can really do.' And I don't know who's going to stop them from doing that."

As for the time line, Church says, "I think there will be multiple clinical infertility solutions involving gene therapy soon."

How soon?

"The next couple years," he says. "It would be very hard to resist." In fact, in August 2017 researchers successfully corrected a common genetic cause of male infertility in mice. As of now, current law in the U.K. would prevent such a procedure in humans.

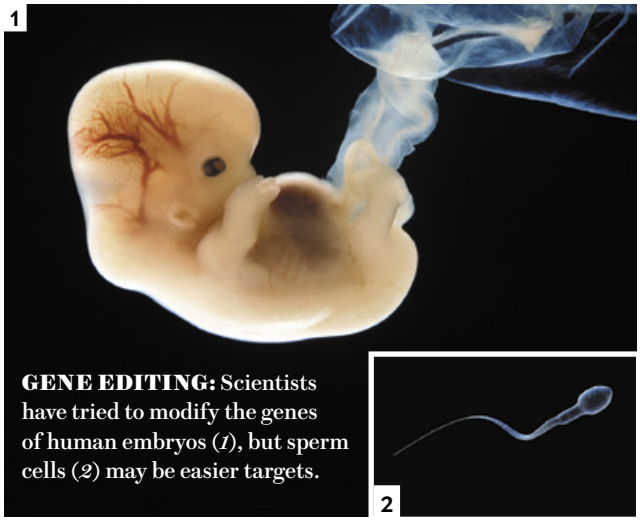
In his talk at the National Academies meeting, Orwig flashed a slide that said: "Germ-Line Gene Therapy Is Technically Feasible Today." Afterward, according to Orwig, a member of the planning committee sidled up to him backstage and said, "Germ-line gene therapy is going to happen, I guarantee it." That sentiment never made it into the final meeting communiqué. But it galvanized Orwig. "Whereas it might have been something I wanted to do hiding under a rock, it's now like, 'You go, man!' Let me get to work and prove to you that I can do it."

In animals, of course.

A GENTLE NUDGE DOWN THE SLOPE

A FEW STEPS DOWN the hallway from Orwig's office is a complex of rooms that houses hundreds of mice. Many of the cages contain what are called nude mice—pink, wrinkled little rodents that resemble scrotums with eyes and feet. They are nude in the sense that they were bred to have compromised immune systems that accommodate cells transplanted from other species—for instance, human spermatogonial stem cells with mutations—to allow researchers to better understand the biology of male infertility.

If, as Church says, "everything is going to be done in animals first," the road to human germ-line modification runs through rooms like these. CRISPR makes the task more efficient ("It is so freaking easy!" Orwig says), but scientists have been able to alter the genes of sperm-making cells for more than two decades, beginning in 1994, when University of Pennsylvania biologist Ralph Brinster (Orwig's mentor) did the pioneering experiments in mice.



GENE EDITING: Scientists have tried to modify the genes of human embryos (*1*), but sperm cells (*2*) may be easier targets.

Male infertility has many causes, including obstructive "plumbing" issues, glitches in the incredibly complex process of sperm creation, and underachieving sperm. But in many cases, males simply can't make sperm at all; the condition, known as nonobstructive azoospermia, affects roughly 350,000 men in the U.S., according to Orwig. Several genes have been associated with the failure to produce sperm, including *tex11* and *sohlh1,* and those cases form the backdrop of the experiment Orwig is eager to do.

What Orwig wants to do is take infertile mice, which have a dysfunctional version of one of these genes, remove the sperm-forming stem cells from their testes and correct the defect in those cells by using the new gene-editing techniques. Once the altered stem cells are grown to sufficient numbers in the test tube and screened for precisely the correct alteration, they can be transplanted back into the testes of the animals. And at least in animal experiments of this kind, there is no need for any fancy molecular tests—if the gene editing is successful, Orwig will know within a couple of months because the infertile males will unambiguously demonstrate their ability to become fathers.

"We've been transplanting stem cells for 25 years in almost every species—mice, rats, hamsters, sheep, goats, pigs, dogs and monkeys," Orwig says. "That's a pretty broad swath of evolution, and in all this time, in all these animals, as far as we know, nothing bad has happened." That is why Orwig is optimistic he can demonstrate that editing the genes of stem cells in mice can reverse infertility.

This may seem like an innocuous animal experiment, but to edit a sperm-forming stem cell is to permanently modify the germ line because the resulting sperm cells pass the correction along to the next generation. A potential treatment for male infertility would cross the red line. And although Orwig has no plans to do the obvious human follow-up in his Pittsburgh lab, a successful preclinical demonstration in mice and primates would provide the impetus for an attempt in the private sector—

NEIL HARDING *Getty Images (1);* GETTY IMAGES (2)

© 2018 Scientific American

which is where Church believes the final steps will unfold. "Sperm-editing efforts will be privately funded," he says, "just like other therapies."

Developing such a clinical treatment would face technical hurdles, of course. For one thing, scientists would have to find a way to maintain human spermatogonial stem cells long enough to select the right ones for transplantation—still not a trivial task. But these male stem cells offer much less of a moving target than embryos, which are dynamic and change rapidly. The Chinese researchers who have attempted gene editing in embryos with CRISPR, for example, have reported both "untoward mutations" and "mosaicism," meaning that some cells in the embryos show successful editing, whereas others do not. Moreover, the DNA of gene-edited stem cells can be screened before an embryo is produced.

That is what makes Orwig's potential mouse experiment so politically inconvenient. Because of prohibitions enacted by Congress in the 1990s, the National Institutes of Health cannot fund any research that involves the destruction of human embryos. A human version of Orwig's proposed mouse experiment might sidestep that prohibition, but it would probably fall under a new obstacle that the House of Representatives introduced two weeks after the December summit on gene editing. In a two-sentence passage buried in the 2,009-page omnibus spending bill of 2015, Congress inserted language forbidding the FDA to consider any medical intervention relying on the use of gene-edited embryos; the wording does not explicitly prohibit editing germ cells, but Stanford University law professor Henry Greely believes "the FDA would take the position that those sperm were more than minimally manipulated human cells that would require FDA approval as a drug or biological product." The regulatory piece, he thinks, could add a decade or two to Church's time line.

That does not mean Orwig's mouse experiment would be against the law—just a gentle nudge down the slippery slope toward germ-line modification. The step across the red line could happen in private IVF clinics, which have a long (and blemished) history of pushing the envelope on new reproductive technologies. "It's such an easy technology to apply that it would only take somebody with a little chutzpah to get together with someone in an IVF clinic and, you know, take a shot at it," says George Daley, a stem cell biologist at Boston Children's Hospital. "This is coming down the pike, and people need to start thinking about it," he notes.

It probably will not happen in the U.S. unless the public—and political—perception of germ-line modification becomes more accommodating, but Orwig is quietly preparing for that day. "We're going to work real hard behind the scenes," he says, "until the worldview changes."

CROSSING BORDERS

THE "WORLDVIEW" on germ-line editing is complicated and contradictory. A majority of Americans do not like the idea of editing genes in either embryos or germ cells, according to a recent analysis of 17 public opinion polls published in the *New England Journal of Medicine*. Yet paradoxically, most people support gene editing in adults "aimed at preventing one's children from inheriting certain diseases." (Robert J. Blendon, lead author of the study, says any intervention on the adult side that is positive for the next generation, including germ cells, would have "considerable public support.") Moreover, the *NEJM* study also pointed out that many of these public opinion polls pose their questions using language that "might not be scientifically precise." In other words, although the National Academies meeting adjourned in December 2015 with a pledge to continue the public conversation about germ-line editing, it is not clear that the public even understands the terms of that conversation. While public forums struggle to find an effective vocabulary, the science races ahead.

As we spoke in his office in spring 2016, Orwig nodded at a scientific reprint sitting on his desk. "I really, really love this paper," he said. He was referring to research published in the March 3, 2016, issue of *Cell Stem Cell* by a group headed by Qi Zhou of Zhejiang University in China. The experiment basically provided a recipe for the in vitro creation of germ cells.

The researchers showed that they could create sperm-forming stem cells in a dish; with a technique currently used in IVF clinics, these cells could be injected into egg cells to create fertile male mice. Harvard's Daley says of this advance: "With the addition of CRISPR, you've got the brave new world."

When Aldous Huxley imagined his brave new world in 1932, the story unfolded under one totalitarian regime, with neither national boundaries nor local regulations. In today's world, germ-line editing in any one place means the germ line is edited everywhere. "Regulation is country-specific, but science crosses borders," Harvard Law's Cohen says. Even if there were laws against germ-line modification in the U.S., you would have to build a wall much higher than the one proposed by Donald Trump to keep American germ lines insulated from an eventual influx of modified DNA.

"If you play out the world to 100 years from now, if anybody does this anywhere, that's the end game," Cohen says. "Over time those people will mate and create offspring and will cross borders and enter our shores. And if the safety and efficacy are worked out, it's inevitable that you're going to have people walking around in the world, and they're going to be reproducing, and they will end up in this country, and those changes will enter the U.S. gene pool."

As I am concluding my visit with Orwig, he glances at the computer on his desk. A reporter has sent an e-mail seeking comment on yet another experiment cozying up to the red line: a group in China has just reported its attempt to edit human embryos (nonviable) to be resistant to HIV infection. "Eventually we'll learn a vocabulary that acknowledges that we're there," Orwig says. "But I feel we're already there." SA

Stephen S. Hall is an award-winning science writer. He has authored six books, including histories of recombinant DNA and cancer immunotherapy, and teaches science communication at New York University.

MORE TO EXPLORE

The Pandora's Box Congress. Michael Rogers in *Rolling Stone;* June 19, 1975.
CRISPR Germline Engineering—The Community Speaks. Katrine S. Bosley et al. in *Nature Biotechnology,* Vol. 33, pages 478–486; May 2015.
Experimental Methods to Preserve Male Fertility and Treat Male Factor Infertility. Kathrin Gassei and Kyle E. Orwig in *Fertility and Sterility,* Vol. 105, No. 2, pages 256–266; February 2016.
Correction of a Pathogenic Gene Mutation in Human Embryos. Hong Ma et al. in *Nature,* Vol. 548, pages 413–419; August 24, 2017.

scientificamerican.com/magazine/sa

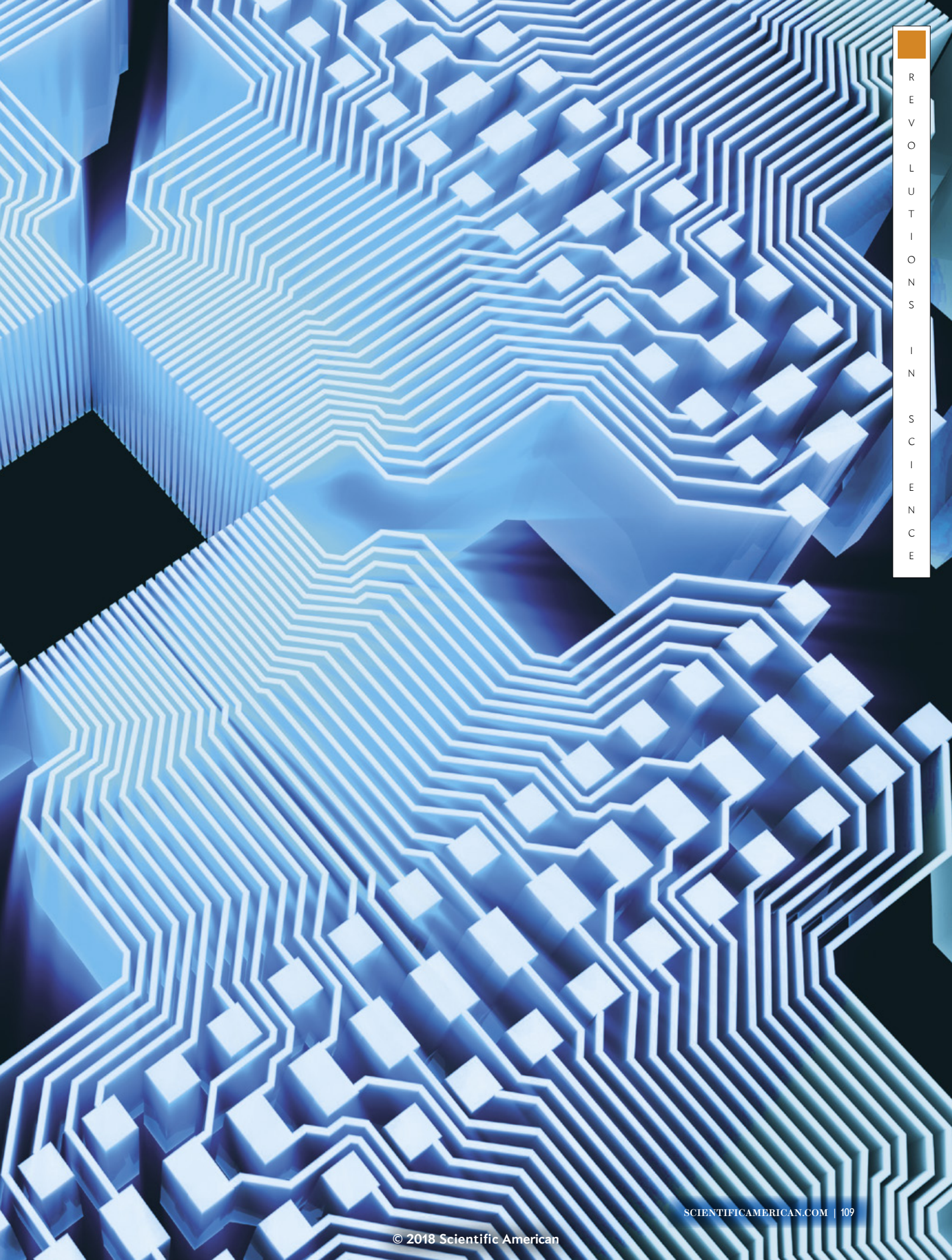
© 2018 Scientific American

QUANTUM COMPUTERS COMPETE FOR "SUPREMACY"

Two technologies may be on the verge of surpassing even the most powerful digital computers in a year or so, but key challenges remain unsolved

By Neil Savage

© 2018 Scientific American



© 2018 Scientific American

SCIENTISTS HAVE LONG DREAMED OF DEVELOPING QUANTUM COMPUTERS, machines that rely on arcane laws of physics to perform tasks far beyond the capability of today's strongest supercomputers. In theory such a machine could create mathematical models too complex for standard computers, vastly extending the range and accuracy of weather forecasts and financial market predictions, among other things. They could simulate physical processes such as photosynthesis, opening new frontiers in green energy. Quantum computing could also jolt artificial intelligence to a vastly higher level of sophistication: If IBM's Watson can already win at *Jeopardy!* and make some medical diagnoses, imagine what an enormously smarter version could do.

But to realize those visions, scientists first have to figure out how to actually build a quantum computer that can perform more than the simplest operations. They are now getting closer than ever, with IBM in November 2017 announcing its most complex quantum system so far and Google saying it is on track to unveil a processor with so-called "quantum supremacy"—capabilities no conventional computer can match.

Small systems exist, but the next steps in the race to make them bigger will have to determine whether quantum computers can deliver on their potential. Scientists and industry players have focused largely on one of two approaches. One cools loops of wire to near –273.15 degrees Celsius, or absolute zero, turning them into superconductors where current flows with virtually no resistance. The other relies on trapped ions—charged atoms of the rare earth element ytterbium held in place in a vacuum chamber by laser beams and manipulated by other lasers. The oscillating charges (in both the wires and the trapped ions) function as quantum bits, or "qubits," which can be harnessed to carry out the computer's operations.

QUANTUM LEAPS

THE TRICK TO EITHER approach is figuring out how to get from already demonstrated systems—containing just a few qubits—to ones that can handle the hundreds or thousands, even millions, required for the kind of heavy lifting that quantum technology seems to promise. In 2016 IBM made a five-qubit quantum processor available to developers, researchers and programmers for experimentation via its cloud portal. The company has made significant progress since then, revealing that it has upgraded its cloud-based quantum computer to a 20-qubit processor—and developed in-house a 50-qubit processor prototype that could be the basis for commercial systems. Both are based on the wire-loop superconducting circuits, as is Google's 72-qubit processor, which the company announced at the annual meeting of the American Physical Society in March 2018. Julian Kelly, an engineer in Google's Quantum Artificial Intelligence Lab, wrote in a blog post that the company is "cautiously optimistic" it can achieve quantum supremacy with its processor.

Those numbers may not seem impressive. But a qubit is much more powerful than the kind of bit that serves as the smallest unit of data in a conventional computer. Those bits are based on the flow of electric current and make up the digital language in which all computing functions: "Off" means 0 and "on" means 1, and those two states encode all of the computer's operations. Qubits, however, are not based on "yes/no" electri-

PRECEDING PAGES: ALFRED PASIEKA *Getty Images*

© 2018 Scientific American

cal switches—but rather on a particle's quantum properties, such as the direction in which an electron spins. And in the quantum world a particle can simultaneously exist in a variety of states more complex than simply on/off—a phenomenon known as superposition. "You can have heads, you can have tails, but you can also have any weighted superposition. You can have 70-30 heads-tails," says Christopher Monroe, a physicist at the University of Maryland, College Park, and founder of IonQ, a start-up working on building a quantum computer with trapped ions. He is currently working on a 32-qubit prototype based on that approach.

The more-than-binary ability to occupy multiple states at once allows qubits to perform many calculations simultaneously, vastly magnifying their computing power. That power grows exponentially with the number of qubits. So at somewhere around 49 or 50 qubits, quantum computers reach the equivalent of about 10 quadrillion bits and become capable of calculations no classical computer could ever match, says John Preskill, a theoretical physicist at the California Institute of Technology. "Whether they will be doing useful things is a different question," he says. Although IBM has reached the 50-qubit mark and Google has passed it, neither has yet demonstrated that they have surpassed the abilities of classical computers.

Both superconducting circuits and trapped ions have a good shot at hitting that fiftyish-qubit threshold, says Jerry Chow, manager of experimental quantum computing at IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y. Conventional thinking would suggest that more qubits means more power—but Chow notes that "it's not just about the number of qubits." He is more focused on the number and quality of calculations the machine can perform, a metric he calls "quantum volume." That includes additional factors such as how fast the qubits can perform the calculations and how well they avoid or correct for errors that can creep in. Some of those factors can work against one another; adding more qubits, for instance, can increase the rate of errors as information passes down the line from one qubit to another. Quantum supremacy is just one possible measure of a quantum computer's capabilities, he says, and the community needs to come up with various tests before it can state that a system has really surpassed classical machines. "As a community we should all be focusing—no matter whether we're working on superconducting qubits or trapped ions or whatever—on pushing this quantum volume higher and higher so we can really make more and more powerful quantum processors and do things that we never thought of," Chow says.

"WHEN YOU MAKE A QUANTUM SYSTEM BIG, IT BECOMES LESS QUANTUM."

BETTER, NOT BIGGER

MONROE COMPARED his five-qubit trapped ion system with IBM's five-qubit processor by running the same simple algorithms on both and found the performance comparable. The biggest difference, he says, is that the trapped ions are all connected to one another via electromagnetic forces: Wiggle one ion in a string of 30 and every other ion reacts, making it easy to quickly and accurately pass information among them. In the wire-loop superconductor circuit only some qubits are connected, which makes passing information a slower process that can introduce errors.

One advantage of superconducting circuits is that they are easy to build using the same processes that make computer chips. They perform a computer's basic logic gate operations—that is, adding, subtracting or otherwise manipulating the bits—in billionths of a second. On the other hand, qubits in this type of system hold their quantum state for only milliseconds—thousandths of a second—so any operation must be completed in that time.

Trapped ions, in contrast, retain their quantum states for many seconds—sometimes even minutes or hours. But the logic gates in such a system run about 1,000 times slower than in superconductor-based quantum computing. That speed reduction probably does not matter in simple operations with just a few qubits, Monroe says. But it could become a problem for getting an answer in a reasonable amount of time as the number of qubits increases. For superconducting qubits, rising numbers may mean a struggle to connect them together.

And increasing the number of qubits, no matter what technology they are used with, makes it harder to connect and manipulate them—because that must be done while keeping them isolated from the rest of the world so they will maintain their quantum states. The more atoms or electrons are grouped together in large numbers, the more the rules of classical physics take over—and the less significant the quantum properties of the individual atoms become to how the whole system behaves. "When you make a quantum system big, it becomes less quantum," Monroe says.

Chow thinks quantum computers will become powerful enough to do at least something beyond the capability of classical computers—possibly a simulation in quantum chemistry—within about five years. Monroe says it is reasonable to expect systems containing a few thousand qubits in a decade or so. To some extent, Monroe says, researchers will not know what they will be able to do with such systems until they figure out how to build them.

Preskill, who is 65, says he thinks he will live long enough to see quantum computers have an impact on society in the way the Internet and smartphones have—although he cannot predict exactly what that effect will be. "These quantum systems kind of speak a language that digital systems don't speak," he says. "We know from history that we just don't have the imagination to anticipate where new information technologies can carry us." SA

Neil Savage is a freelance writer based in Lowell, Mass.

MORE TO EXPLORE

Experimental Comparison of Two Quantum Computing Architectures. Norbert M. Linke et al. in *Proceedings of the National Academy of Sciences USA*, Vol. 114, No. 13, pages 3305–3310; March 28, 2017.

Characterizing Quantum Supremacy in Near-Term Devices. Sergio Boixo et al. in *Nature Physics*. Published online April 23, 2018. www.nature.com/articles/s41567-018-0124-x

scientificamerican.com/magazine/sa

© 2018 Scientific American

Sleeping Beauties of Science

Some of the best research can slumber for years

Discovery lies at the core of science, but rediscovery can be just as important. Indiana University Bloomington researchers combed through 22 million scientific papers published over the past century and found, to their surprise, dozens of "Sleeping Beauties"—studies that sat dormant for years and then suddenly got noticed. The top finds—the ones that languished in obscurity the longest and later got the most intense attention from the scientific community—came from chemistry, physics and statistics. What tended to wake up these sleeping findings? Scientists from other disciplines, such as the medical field, in search of fresh insights, as well as the ability to test once theoretical postulations. Qing Ke, a graduate student in informatics who worked on the project, says that most likely Sleeping Beauties will become even more common in the future because of the increasing availability and accessibility of scientific literature.

—Amber Williams

Top 15 Sleeping Beauties
Physics
Chemistry
Statistics
Metallurgy
Citations after "awakening"
300 citations
Number of citations per year
Here H.M.F. Freundlich publishes the first mathematical model of adsorption, when atoms or molecules adhere to a surface. Today both environmental remediation and decontamination in industrial settings rely heavily on adsorption.
John Turkevich and his colleagues' paper explains how to suspend gold nanoparticles in liquid. It owes its awakening to the medical field, which now employs gold nanoparticles to detect tumors and deliver drugs.
2010
2000
1990
1980
1970
1960
1950
1940
1930
1920
1910
1900
1 "Concerning Adsorption in Solutions"
2 "Preparation of Graphitic Oxide"
3 "The Scherrer Formula for X-ray Particle Size Determination"
4 "Wettability of Porous Surfaces"
5 "A Study of the Nucleation and Growth Processes in the Synthesis of Colloidal Gold"
6 "On Lines and Planes of Closest Fit to Systems of Points in Space"
7 "The Tension of Metallic Films Deposited by Electrolysis"
8 "Emulsions"
9 "Resistance of Solid Surfaces to Wetting by Water"
10 "Probable Inference, the Law of Succession, and Statistical Inference"
11 "The Constitution and Fundamental Properties of Solids and Liquids, Part I: Solids"
12 "Note on an Approximation Treatment for Many-Electron Systems"
13 "Relations between the Elastic Moduli and the Plastic Properties of Polycrystalline Pure Metals"
14 "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?"
15 "The Dynamics of Capillary Flow"
This paper introduces Hummer's method, a technique for making graphite oxide. The recent fervor over graphene's potential has brought the 1958 paper into the limelight. Graphite oxide could serve as a reliable intermediate for the 2-D material.
From publication to "awakening," little to no activity
Albert Einstein and his colleagues' famous thought experiment in quantum physics—now known as the EPR paradox—was discussed theoretically when it first came out. But it was not until the 1970s that physicists had the experimental means to test quantum entanglement.

SOURCE: "DEFINING AND IDENTIFYING SLEEPING BEAUTIES IN SCIENCE," BY QING KE ET AL., IN *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES USA*, VOL. 112, NO. 24; JUNE 16, 2015

Graphic by Tiffany Farrant-Gonzalez

© 2018 Scientific American

BRIGHT HORIZONS 36

SCIENTIFIC AMERICAN Travel

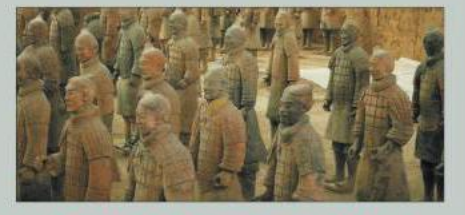
China | October 27th – November 6th, 2018

Embark on Bright Horizons 36's journey of discovery in China and send your senses reeling. Taste dim sum in Hong Kong and local brewed beer in Qingdao. Soak in powder blue hot springs and taste the umami of oyster pancakes in the Taiwanese night markets. Hear the roar of urban life and share the quiet of China's temples and gardens with a friend.

While we sail the Yellow Sea and the East and South China Seas, absorb Earth science from an expert and study the roots of Chinese culture in exclusive Bright Horizons classes.

The time is ripe to visit China and Taiwan. Update your Earth science knowledge and get an unparalleled view of China and her history and culture.

Cruise prices vary from $3,599 for an Ocean View Stateroom to $9,624 for a Owner's Suite, per person (pp) based on double occupancy. For those attending our SEMINARS, there is a $1,675 fee. Add'l pp fees: gov't taxes and fees ($440), a non-refundable Booking Service Fee ($150), and the Tour Leader gratuities ($100), and the option Program/Conference fee ($1,675). The Program, cruise pricing, and options are subject to change. For more information email us at info@InsightCruises.com.



ULTIMATE CHINA

PRE-CRUISE LAND TRIP OCTOBER 19TH – 27TH, 2018

Enhance your once-in-a lifetime trip to China with an optional 8-night, 5-star precruise tour of Beijing, Xi'an's Terracotta Warriors, and rural Guilin's iconic landscapes with Insight Cruises and Imperial Tours. You'll enjoy privileged access behind the scenes in the Forbidden City's "Palace of Double Glory"; see the Terracotta Warriors as visiting dignitaries do, walking between and among them; view 1,000 y.o. Tang frescos up close. Dine *on* the Great Wall, raft through rural China, visit a distinguished art collector — immerse yourself in China's ways to enrich your Bright Horizons cruise conference.

If you have wanted to see China's key sites but hesitate because of language, unfamiliarity with available housing, intra-China travel arrangements (this package includes three flights), or analysis paralysis, we can take you there. Effortlessly. Accompanied by a China host as well as very select area guides, you will see China's touchstones in a seamless and unforgettable experience.

Insight Cruises and Imperial Tours will take care of all the details so you can focus on China with peace of mind.

SEMINARS

EARTH SCIENCE:

- Scales of Change: Why Time Matters to Science and Being Human
- Climate Through the Ages
- How Mountain Building and Climate Shape the Land
- The Early Earth and Life's Origins
- Life: A Brief History
- The "Ice Ages", Rapid Climate Change, and Human Impacts

SPEAKER: Ron Martin, Ph.D.

UNDERSTANDING CONTEMPORARY CHINA:

- How History Informs Contemporary Chinese Attitudes and Actions
- China as a Global Power

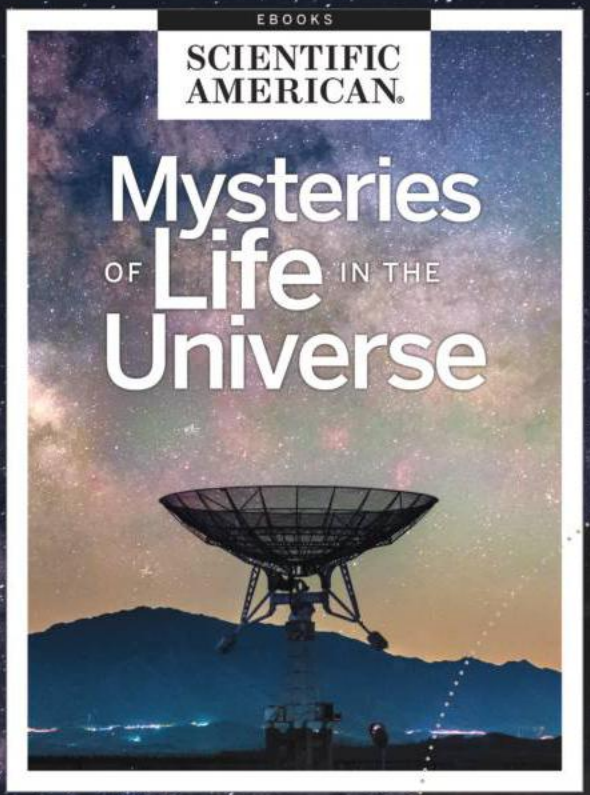
EXPERIENCING IMPERIAL CHINA:

- The Grand Canal and the Great Wall
- Being a Chinese Peasant
- Being a Chinese Scholar
- Imperial China and the Maritime World

SPEAKER: Andrew Wilson, Ph.D.

FOR MORE INFO Please email: **info@InsightCruises.com** or visit: **ScientificAmerican.com/travel**

CST# 2065380-40 °Scientific American, Inc.



Examine the Big Questions

How did life begin?
Are we alone in the universe?
In this eBook, we explore several theories on the origin of life, some of its extreme forms and the ongoing search for signs – or sentience – on distant worlds.

sciam.com/ebooks

SCIENTIFIC AMERICAN eBooks